The Pennsylvania State University
APPLIED RESEARCH LABORATORY
P.O. Box 30
State College, PA 16804

HIGH ORDER NONLINEAR ESTIMATION WITH SIGNAL PROCESSING APPLICATIONS

by

T. W. Hilands

Technical Report No. TR 93-08
August 1993

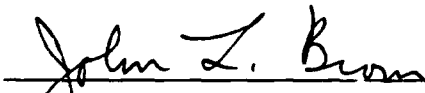
L.R. Hettche, Director
Applied Research Laboratory

We approve the thesis of Thomas W. Hilands.

Date of Signature

Oct. 6, 1992

Stelios C. A. Thomopoulos
Associate Professor of Electrical Engineering
Thesis Advisor
Chair of Committee

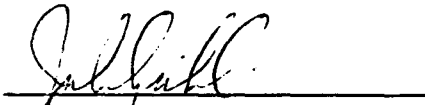


Oct. 6, 1992.

John L. Brown
Professor Emeritus of Electrical Engineering

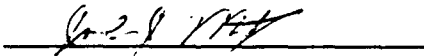
10/6/92

Frank W. Symons
Senior Research Associate



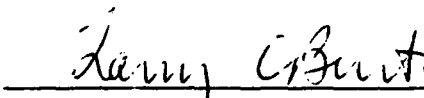
10/6/92

John E. Dzielski
Research Associate



10/6/92

John J. Metzner
Professor of Computer Engineering



10/6/92

Larry C. Burton
Professor of Electrical and Computer Engineering
Head of the Department of Electrical and Computer Engineering

Abstract

Two high order vector filters (HOFs) are developed for estimation in non-Gaussian noise. These filters are constructed using nonlinear functions of the innovations process. They are completely general in that the initial state covariance, the measurement noise covariance, and the process noise covariance can all have non-Gaussian distributions. The first filter is designed for systems with asymmetric probability densities. The second is designed for systems with symmetric probability densities. Experimental evaluation for estimation in non-Gaussian noise, formed from Gaussian sum distributions, shows that these filters perform much better than the standard Kalman filter, and close to the optimal Bayesian estimator.

The problem of high resolution parameter estimation of superimposed sinusoids is addressed using nonlinear filtering techniques. Six separate nonlinear filters are evaluated for the estimation of the parameters of sinusoids in white and colored Gaussian noise. Experimental evaluation demonstrates that the nonlinear filters perform close to the Cramer-Rao bound for reasonable values of the initial estimation error. The recursive technique developed here is well suited for time-varying systems and for measurements with short data lengths.

A general approach to model order selection is presented based on joint detection/estimation theory. The approach involves the simultaneous application of maximum *a posteriori* (MAP) detection and nonlinear estimation using either the extended Kalman filter when the noise is Gaussian, or the extended high order filter (EHOF) when the noise is in non-Gaussian. The problem is formulated as a multiple hypothesis testing problem with assumed known *a priori* probabilities for each hypothesis. Experimental evaluation of the approach demonstrates excellent perfor-

mance in selecting the correct model order and estimating the system parameters for SNR's as low as -5 dB.

A nonlinear adaptive detector/estimator (NADE) is introduced for single and multiple sensor data processing. The problem of target detection from returns of monostatic sensor(s) is formulated as a nonlinear joint detection/estimation problem on the unknown parameters in the signal return. The unknown parameters involve the presence of the target, its range, azimuth, and Doppler velocity. The problems of detecting the target and estimating its parameters are considered jointly. A bank of spatially and temporally localized nonlinear filters is used to estimate the a *posteriori* likelihood of the existence of the target in a given space-time resolution cell. Within a given cell, the localized filters are used to produce refined spatial estimates of the target parameters. Excellent performance is obtained using this technique for single sensor processing and for centralized data fusion.

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

List of Figures

List of Figures (continued)

List of Tables

List of Tables (continued)

Chapter 1

Introduction

In this thesis a new high order filter (HOF) is developed for estimation in non-Gaussian noise. It is shown that this new filter yields improved performance over the standard linear Kalman filter and is less computationally intensive than optimal non-Gaussian filtering techniques such as Gaussian sum filters. This thesis also addresses parameter estimation in the context of several signal processing problems. These problems, which are formulated as nonlinear estimation problems, have been traditionally addressed using other parametric and nonparametric techniques. It is shown that nonlinear filtering techniques, including the nonlinear version of the HOF, designated the extended high order filter (EHOF), can perform very well for estimation of signal parameters in Gaussian and non-Gaussian noise.

1.1 Motivation for the Study

The standard Kalman filter does not use the higher moments of the density functions and therefore cannot adequately deal with non-Gaussian distributions. Many of the existing techniques for estimation in the presence of non-Gaussian noise require accurate knowledge of the density functions. Given this knowledge, they attempt to approximate these functions using Gaussian sums or other approximations address the problem of nonlinear estimation in non-Gaussian noise. Other methods make simplifying assumptions such as symmetrical distributions, small plant noise, or small measurement noise in order to develop approximate filters. This motivated a study of the filtering problem from a more general point of view. The goal of this study is to develop filtering algorithms for systems in non-Gaussian noise that

use knowledge of the moments of the *a priori* distributions. In these algorithms no assumptions are made about the power of the noise or the shape of the probability density function.

Several specific problems in the signal processing area are of interest in the application of nonlinear parameter estimation techniques in the presence of Gaussian and non-Gaussian noise. These problems are also associated with estimating the parameters of sinusoids.

A problem that has attracted a large amount of research is that of harmonic retrieval. This problem consists of estimating some or all of the frequencies, amplitudes, damping coefficients, and phases of superimposed sinusoids in white or colored, Gaussian or non-Gaussian noise. Much of the work in the area of high resolution spectral estimation or harmonic retrieval has been based on fitting an autoregressive (AR) or autoregressive moving average (ARMA) model to the received data. However, the performance of most modern high resolution estimation techniques is severely degraded at low SNR's and/or short data lengths. This is probably due to the fact that these techniques are heuristic least squares modifications of algorithms that yield exact results when there is no noise or when the available data is infinite. Quite often the initial conditions on a problem can be bounded so that fairly accurate *a priori* estimates can be obtained. The harmonic retrieval problem is successfully addressed in this thesis with nonlinear estimation techniques.

A separate but related problem is that of model order selection. The objective in model order selection is to determine the number of sinusoids embedded in Gaussian and non-Gaussian noise. This problem is approached in this thesis with

joint detection/estimation techniques.

The joint detection/estimation (JD/E) procedure is presented in Chapter 5. The procedure is structured mathematically so that it can be employed against problems with model uncertainty, initial condition uncertainty, or both. The JD/E technique can be applied to any type of noise, assuming the density function is known. This technique is applied in subsequent chapters for selected sinusoidal detection and parameter estimation problems.

Joint detection/estimation techniques can also be applied to the estimation of Doppler shift and time delay from an echo of a transmitted signal. Traditional solutions for this problem are based on Fourier transform implementations and generally have poor resolution in the presence of short data lengths. It is shown how estimates from multiple sensors can be *combined* to form improved estimates of target range, geometric angle, and velocity.

1.2 Scope of the Thesis

Chapter 2 discusses the fundamentals of estimation theory and presents the primary techniques currently used to perform nonlinear estimation in Gaussian noise, and linear estimation in non-Gaussian noise. This chapter is essentially composed of background material that is needed for an understanding of the remainder of the thesis.

Chapter 3 presents a general solution to the problem of estimation in the presence of non-Gaussian noise. The solution is based on high order powers of the innovations process. The solution is entirely general in that the plant noise, the measurement noise, or the initial estimation error can be non-Gaussian with

symmetrical or asymmetrical distributions. The performance of the filter for non-Gaussian noise is compared to exact Bayesian filters. Non-Gaussian distributions are created using a sum of Gaussian distributions. Bayesian filters can be constructed to give optimal performance for Gaussian sum distributions. The intent of this comparison is to numerically evaluate the performance of the non-Gaussian filters and to determine where these filters provide improvement in state estimation over the standard Kalman filter. It is shown that the high order filter (HOF) performs better than the standard Kalman filter, but not quite as well as the optimal Gaussian sum filter.

Chapter 4 shows that nonlinear filtering techniques can be used for high resolution harmonic retrieval. Traditional approaches in this area have been concerned with Fourier transforms or techniques based on autorecursive (AR) or autorecursive moving average (ARMA) estimation. Many of these approaches are batch estimators and, as such, cannot adequately deal with time varying systems. In addition, most of these techniques cannot take advantage of *a priori* estimates of the initial system state. It is shown that nonlinear filtering methods can give highly accurate estimates (approaching the CR bound) of the parameters of sinusoids in white and colored Gaussian noise. A particularly attractive filter to use in the harmonic retrieval problem is the minimum variance filter. This filter requires exact expressions for expected values of nonlinear functions of the state variables during each iteration of the filter equations. Closed form expressions for these expected values are developed for the specific nonlinear functions used in the harmonic retrieval problem. Using these expressions it is expected that the minimum variance filter should give better state estimates than the extended Kalman filter (EKF) especially when there are large errors in the initial estimates. In this chapter Monte Carlo simu-

lations are used to compare the performance of several nonlinear filters to the CR bound. Studies are performed to determine the effect of poor initial conditions on the performance of these nonlinear filters.

The joint detection/estimation (JD/E) procedure is presented in Chapter 5. The procedure is structured mathematically so that it can be employed against problems with model uncertainty, initial condition uncertainty, or both. The JD/E technique can be applied to any type of noise, assuming the density function is known. This technique is applied in subsequent chapters for selected sinusoidal detection and parameter estimation problems.

The JD/E technique is used in Chapter 6 to perform model order selection. A general approach is presented for determining the number of sinusoids present in measurements corrupted by additive white Gaussian and non-Gaussian noise. Experimental evaluation of this approach demonstrates excellent performance for model order selection and system parameter estimation in both Gaussian and non-Gaussian noise.

Chapter 7 uses the JD/E approach to estimate time delay and Doppler shift from echos of a transmitted waveform. The problem of target detection from returns of monostatic sensor(s) is formulated as a nonlinear joint detection/estimation problem on the unknown parameters in the signal return. In this chapter it is assumed that the target has been detected. The JD/E procedure is applied by segmenting a large initial estimation error into smaller regions of uncertainty and operating an independent nonlinear filter to perform parameter estimation for each of these regions. It is found that this approach can help solve the problem of convergence to local minima, which is characteristic of estimators such as the EKF.

In Chapter 8, the problems of detecting the target and estimating its parameters are considered jointly. The fusion of parameter estimates from two spatially separated sensors is accomplished using the JD/E approach. Several hypotheses are postulated for detection. Each hypothesis corresponds to the ability of each sensor to detect the target in its area of coverage. The *a priori* probabilities of each decision are based on the area of coverage of the two sensors. For each hypothesis, a nonlinear filter recursively estimates target parameters. The maximum likelihood estimate for a given hypothesis is then determined as a weighted sum of the estimates from each of the local hypotheses, with the *a posteriori* probability being used as the weighting function. It is shown experimentally that excellent performance can be obtained for both target detection and target parameter estimation using this technique.

Chapter 2

Optimal and Suboptimal Estimation

The purpose of this chapter is to briefly cover the fundamentals of estimation theory and to discuss several techniques for nonlinear estimation found in existing literature. Section 2.1 presents the basic concepts of estimation theory and some of the properties of estimators. Section 2.2 presents optimal Bayesian estimation. This section also presents the derivation of the linear Kalman filter, which is the optimal estimator for linear systems in additive white Gaussian noise. In Section 2.3 Bayesian approximations are discussed. These approximations entail methods for estimation of the *a posteriori* density function. Section 2.4 discusses nonlinear filtering techniques for nonlinear systems in additive Gaussian noise. Section 2.5 presents techniques for linear filtering in non-Gaussian noise. The overall goal of this chapter is to explain the basics of estimation theory and to show the evolution of the optimal linear estimator, the Kalman Filter, into techniques for nonlinear estimation. This will lay the groundwork for further discussions on new techniques presented in this thesis for suboptimal estimation in non-Gaussian noise and for applications of nonlinear filtering to specific signal processing problems. In this thesis only discrete time (i.e. sampled data) estimation problems are addressed.

2.1 Fundamentals of Estimation Theory

Estimation theory addresses the process of determining the value of some uncertain quantity based on available pertinent information. Consider the problem of estimating the n-dimensional time invariant parameter vector $\mathbf{x}$ from observations represented by the m-dimensional vector $\mathbf{z}_k$. The measurements are described by

the nonlinear relation

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}, k, \mathbf{v}_k).$$

where $\mathbf{v}_k$ is random noise. The estimate $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = \mathbf{e}_k(\mathbf{Z}_k)$$

where $\mathbf{Z}_k$ is the set of all measurements $(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_k)$. The function $\mathbf{e}_k$ is called the estimator of $\mathbf{x}$. There are two basic models for the parameter $\mathbf{x}$:

(1) *Nonrandom*, when $\mathbf{x}$ has an unknown deterministic value.

(2) *Random*, when the parameter $\mathbf{x}$ has a *priori* probability density function (PDF) $p(\mathbf{x})$.

For nonrandom parameters it is desired that the estimates converge to the true value as $k \rightarrow \infty$. For random time invariant parameters, a realization of $\mathbf{x}$ is drawn from a population with the assumed PDF. One would like each measurement to yield an estimate that converges in some well-defined probabilistic sense independent of the particular realization of $\mathbf{x}$.

Optimal estimation defines the best estimate of a parameter based on some well-chosen criteria of optimality. Since different criteria may lead to different optimal estimates for the same quantity, one may settle for feasible or acceptable estimates according the following rules [1]:

(1) An estimate $\hat{\mathbf{x}}$ is unbiased if it satisfies the relation

$$E[\hat{\mathbf{x}}] = E[\mathbf{x}]$$

(2) An estimate $\hat{\mathbf{x}}$ is a consistent estimate if it converges in probability to $\mathbf{x}$, i.e.

$$\lim_{k\to\infty} \text{prob}[||\hat{\mathbf{x}} - \mathbf{x}|| \geq \epsilon] = 0 \quad \text{for arbitrarily small } \epsilon$$

A consistent estimate is always unbiased.

(3) An efficient estimate $\hat{\mathbf{x}}$ is the unbiased estimate of $\mathbf{x}$ with the minimum variance, i.e.,

$$\sigma_{\hat{\mathbf{x}}}^2 = E[||\hat{\mathbf{x}} - \mathbf{x}||]^2 \leq E[||\mathbf{y} - \mathbf{x}||^2] = \sigma_{\mathbf{y}}^2$$

for all other estimates $\mathbf{y}$ of $\mathbf{x}$.

(4) An estimate $\hat{\mathbf{x}}$ is called sufficient if it contains all of the information in the set of observed values regarding the parameter $\mathbf{x}$ to be evaluated. Any statistic related to a sufficient estimate is called a sufficient statistic.

Several estimation techniques have been used for the estimation of random parameters. Many of these techniques are derived from or related to Bayesian estimation.

The maximum a *posteriori* (MAP) estimate is obtained by maximizing the conditional density

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{z})}$$

with respect to the unknown parameter vector $\mathbf{x}$. Since $p(\mathbf{z})$ is not a function of $\mathbf{x}$, the MAP estimate may be obtained by maximizing the joint density

$$p(\mathbf{z}|\mathbf{x})\, p(\mathbf{x}) = p(\mathbf{z}, \mathbf{x})$$

with respect to $\mathbf{x}$. This can be accomplished by maximizing the natural logarithm

of this quantity so that the MAP estimate can be expressed by

$$\left.\frac{\partial \ln p(\mathbf{z},\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}(\mathbf{z})} = \left.\frac{\partial \ln p(\mathbf{z}|\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}(\mathbf{z})} + \left.\frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}(\mathbf{z})} = 0.$$

In the case where $p(\mathbf{x})$ is unknown the best choice of $\mathbf{x}$ is made based on maximizing the likelihood function $p(\mathbf{z}|\mathbf{x})$. The maximum likelihood (ML) estimate is given by

$$\left.\frac{\partial \ln p(\mathbf{z}|\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}(\mathbf{z})} = 0.$$

It is clear that the ML estimate is inferior to the MAP estimate since it does not consider prior information about the random vector $\mathbf{x}$. However, the ML estimate may be useful in situations where: (1) the parameter $\mathbf{x}$ is unknown but not random, (2) the *a priori* density of $\mathbf{x}$ is unknown, or (3) the density functions $p(\mathbf{x}|\mathbf{z})$ or $p(\mathbf{x},\mathbf{z})$ are more difficult to compute than $p(\mathbf{z}|\mathbf{x})$.

Consider the problem of estimating a nonrandom parameter vector $\mathbf{x}$ from a single linear measurement of this vector in Gaussian noise. In this case the measurement model is given by

$$\mathbf{z} = H\mathbf{x} + \mathbf{v}$$

where $\mathbf{v} \sim N(0, R)$. The likelihood function is given by

$$p(\mathbf{z}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{1}{2}}\,|R|^{\frac{1}{2}}}\exp(-\frac{1}{2}(\mathbf{z}-H\mathbf{x}))^T R^{-1}(\mathbf{z}-H\mathbf{x})),$$

and the maximum likelihood estimate of $\mathbf{x}$ is the root of the equation

$$-2(\mathbf{z}-H\mathbf{x})^T\, R^{-1}\, H = 0$$

leading to the ML estimate

$$\hat{\mathbf{x}} = (H^T R^{-1} H)^{-1} H^T R^{-1} \mathbf{z}.$$

Least squares estimates are obtained by minimizing the sum of the squared error between the measurements and the measurement model. It can be shown [2] that if the noises are independent, identically distributed (i.i.d.), zero-mean, Gaussian random variables the least squares estimate is the same as the ML estimate.

The minimum mean square error (MMSE) estimate, or minimum variance estimate, is obtained by minimizing the expected value of the mean square error $E[(\hat{\mathbf{x}} - \mathbf{x})^2 | \mathbf{Z}_k]$ of the estimate based on the data up to and including time k. The solution is the conditional mean expressed in terms of the conditional PDF $\hat{\mathbf{x}} = \int \mathbf{x} p(\mathbf{x}|\mathbf{Z}_k) d\mathbf{x}$.

2.2 Optimal Bayesian Estimation

An optimal estimate is defined as the minimum variance estimate or the mean of the conditional density function. It will be shown in this section that recursion relations can be set up to determine the conditional density based on Bayes' rule.

Consider the problem of estimating a time varying n-dimensional state vector $\mathbf{x}_k$, where the state evolves according to the plant equation

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{w}_k). \tag{2.1}$$

The state $\mathbf{x}_k$ is observed through the m-dimensional measurement vector $\mathbf{z}_k$ given by

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{v}_k) \tag{2.2}$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are mutually independent white noise sequences. The problem is to estimate the state $\mathbf{x}_k$ from the measurements $\mathbf{Z}_k$, where $\mathbf{Z}_k$ is the set of all measurements $(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_k)$. The objective of Bayesian parameter estimation is to recursively calculate the a *posteriori* density of the state. This density, also referred to as the filtering density, can be obtained [3] through the recursion relations

$$p(\mathbf{x}_k|\mathbf{Z}_k) = \frac{p(\mathbf{x}_k|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\mathbf{x}_k)}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \tag{2.3}$$

$$p(\mathbf{x}_k|\mathbf{Z}_{k-1}) = \int p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})\, p(\mathbf{x}_k|\mathbf{x}_{k-1}) d\mathbf{x}_{k-1} \tag{2.4}$$

where

$$p(\mathbf{z}_k|\mathbf{Z}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\mathbf{x}_k) d\mathbf{x}_k. \tag{2.5}$$

The initial density $p(\mathbf{x}_0|\mathbf{z}_0)$ is given by

$$p(\mathbf{x}_0|\mathbf{z}_0) = \frac{p(\mathbf{z}_0|\mathbf{x}_0)\, p(\mathbf{x}_0)}{p(\mathbf{z}_0)}. \tag{2.6}$$

The density $p(\mathbf{z}_k|\mathbf{x}_k)$ in equation (2.3) can be determined by the a *priori* measurement noise density $p(\mathbf{v}_k)$ and the measurement equation (2.2). Likewise, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ in (2.4) is determined from $p(\mathbf{w}_{k-1})$ and equation (2.1). Knowledge of these densities and $p(\mathbf{x}_0)$ determines $p(\mathbf{x}_k|\mathbf{Z}_k)$ for all k. However, the major difficulty with recursive Bayesian estimation is the closed form solution of the integration in (2.4). This integral can be solved only for linear state and measurement equations with Gaussian statistics, and a limited set of nonlinear systems.

The advantage of using Bayesian estimation is that once the a *posteriori* density is obtained one can compute estimates based on any estimation criteria.

For example, the most probable estimate is found by maximizing the probability that $(\hat{\mathbf{x}}_{k|k} = \mathbf{x})$, yielding the solution $\hat{\mathbf{x}}_{k|k} = \text{Mode}\{p(\mathbf{x}_k|\mathbf{Z}_k)\}$. When the *a priori* density is uniform, this estimate is identical to the ML estimate. If the criteria is to minimize $\int ||\mathbf{x}_k - \mathbf{x}||^2 p(\mathbf{x}_k|\mathbf{Z}_k)$, the solution is $\hat{x}_{k|k} = E[\mathbf{x}_k|\mathbf{Z}_k]$. This is the conditional mean estimate. If the criteria is to minimize the maximum of $|\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}|$, the solution is the minimax estimate defined by $\mathbf{x}_{k|k} = \text{Median}\{p(\mathbf{x}_k|\mathbf{Z}_k)\}$.

In the case of linear systems in Gaussian noise, equations (2.3 - 2.6) can be evaluated and the a *posteriori* density is Gaussian for all k. The conditional mean and covariances for this system are the Kalman filter equations, which were first introduced by R. E. Kalman [4]. In the development to follow the Kalman filter relations are derived from the Bayesian recursion formulas. This derivation is based on a similar development by Ho and Lee [5]. The linear plant and measurement models have the form

$$\mathbf{x}_k = \Phi_{k-1}\mathbf{x}_{k-1} + \Gamma_{k-1}\mathbf{w}_{k-1}$$

$$\mathbf{z}_k = H_k\mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent, white, Gaussian sequences with

$$\begin{aligned} E[\mathbf{v}_k] = E[\mathbf{w}_k] = 0 \quad \forall\, k \\ E[\mathbf{v}_k\mathbf{v}_j^T] = R_k\delta_{kj};\;\; E[\mathbf{w}_k\mathbf{w}_j^T] = Q_k\delta_{kj};\;\; E[\mathbf{v}_k\mathbf{w}_j^T] = 0 \;\; \forall\, j,k \end{aligned} \tag{2.7}$$

Starting with the initial conditions that $p(\mathbf{x}_0|\mathbf{z}_0)$ is Gaussian and

$$\begin{aligned} E[(\mathbf{x}_0|\mathbf{z}_0) = \hat{\mathbf{x}}_{0|0} \\ Cov[\mathbf{x}_0|\mathbf{z}_0] = P_{0|0}. \end{aligned} \tag{2.8}$$

From (2.7) it is noted that $p(\mathbf{x}_k|\mathbf{Z}_{k-1})$ is Gaussian and independent of $\mathbf{v}_k$ so

that

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &= E[\mathbf{x}_k|\mathbf{Z}_{k-1}] = \Phi_{k-1}\hat{\mathbf{x}}_{k-1|k-1} \\ P_{k|k-1} &= Cov[\mathbf{x}_k|\mathbf{Z}_{k-1}] = \Phi_{k-1}P_{k-1|k-1}\Phi_{k-1}^T + \Gamma_{k-1}Q_{k-1}\Gamma_{k-1}^T.\end{aligned} \tag{2.9}$$

Similarly, $p(\mathbf{z}_k|\mathbf{Z}_{k-1})$ is Gaussian and

$$\begin{aligned}E[\mathbf{z}_k|\mathbf{Z}_{k-1}] &= H_k\Phi_{k-1}\hat{\mathbf{x}}_{k-1|k-1} \\ Cov[\mathbf{z}_k|\mathbf{Z}_{k-1}] &= H_kP_{k|k-1}H_k^T + R_k\end{aligned} \tag{2.10}$$

Finally $p(\mathbf{z}_k|\mathbf{x}_k)$ is Gaussian with

$$\begin{aligned}E[\mathbf{z}_k|\mathbf{x}_k] &= H_k\mathbf{x}_k \\ Cov[\mathbf{z}_k|\mathbf{x}_k] &= R_k.\end{aligned} \tag{2.11}$$

Using (2.9 - 2.11) in (2.3) gives

$$\begin{aligned}p(\mathbf{x}_k|\mathbf{Z}_k) = &\frac{|H_kP_{k|k-1}H_k^T + R_k|^{1/2}}{(2\pi)^{n/2}|R_k|^{1/2}|P_{k|k-1}|^{1/2}} \\ &\times \exp\{-\frac{1}{2}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^TP_{k|k-1}^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\ &+ (\mathbf{z}_k - H_k\mathbf{x}_k)^TR_k^{-1}(\mathbf{z}_k - H_k\mathbf{x}_k) \\ &+ (\mathbf{z}_k - H_k\hat{\mathbf{x}}_{k|k-1})^T(H_kP_{k|k-1}H_k + R_k)^{-1}(\mathbf{z}_k - H_k\hat{\mathbf{x}}_{k|k-1})]\}.\end{aligned}$$

Completing the square in the exponent gives

$$\begin{aligned}p(\mathbf{x}_k|\mathbf{Z}_k) = &\frac{|H_kP_{k|k-1}H_k^T + R_k|^{1/2}}{(2\pi)^{n/2}|R_k|^{1/2}|P_{k|k-1}|^{1/2}} \\ &\times \exp\{-\frac{1}{2}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^TP_{k|k}^{-1}(\mathbf{x}_k - \hat{x}_{k|k})\}\end{aligned} \tag{2.12}$$

where

$$\begin{aligned}\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + K_k\tilde{\mathbf{z}}_k \\ P_{k|k} &= (I_n - K_kH_k)P_{k|k-1}\end{aligned} \tag{2.13}$$

K_k is called the Kalman gain and is given by

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1}, \tag{2.14}$$

and the innovations $\tilde{\mathbf{z}}_k$ are defined by

$$\tilde{\mathbf{z}}_k = (\mathbf{z}_k - H_k \hat{\mathbf{x}}_{k|k-1}). \tag{2.15}$$

The filter error covariance in (2.13) can also be expressed as

$$P_{k|k}^{-1} = P_{k|k-1}^{-1} + H_k R_k^{-1} H_k \tag{2.16}$$

where I_n is the n-dimensional indentity matrix. Since the *a posteriori* density is Gaussian, $\hat{\mathbf{x}}_{k|k}$ is the most probable, the conditional mean, and the minimax estimate.

Equations (2.9) and (2.13) constitute the Kalman filter equations. These equations give the optimal, or minimum variance, estimator for linear systems in additive white Gaussian noise. Equation (2.9) is used to extrapolate or predict the estimate from time $k-1$ to time k based on the plant characteristics. Equation (2.13) updates or filters this estimate at time k based on the measurement. An important note about the filter equation (2.13) is that the filtered estimate $\hat{\mathbf{x}}_{k|k}$ is a linear function of the innovations for the optimal linear filter. It will be shown in Chapter 3 how higher order powers of the innovations can be used to develop filters for linear systems in non-Gaussian noise with symmetrical and asymmetrical probability density functions.

The Kalman filter equations can be derived in many ways. Gelb [6] uses the matrix minimum principle on the *a posteriori* variance to obtain the Kalman

filter relations. Kronhamn [7] derives the filter by geometrically demonstrating the orthogonality of the estimation error to the measurement error. Chui and Chen [8] use stochastic operator theory. Jazwinski [9] uses stochastic calculus to come up with the relations for continuous-time systems. Kailath [16] derives the filter using the innovations method. Using this technique the observed process is first converted to a white noise process by means of a causal invertable linear transformation. The problem then becomes one of parameter estimation in white noise. The solution to this simplified problem can then be expressed in terms of the original observations by means of the inverse of the original whitening filter.

Although the Kalman filter is an optimal estimator for linear systems in Gaussian noise, its performance for nonlinear models and non-Gaussian noise is highly dependent on the degree of nonlinearity or non-Gaussianity in the plant and measurement equations. Nonlinear models are generally treated with the extended Kalman filter in which the state and measurement models are linearized about the most recent estimate. This method generally works well in low noise environments. In large noise environments, where the estimation error is large, the Taylor series expansions can be very inaccurate [6].

Filtering in non-Gaussian noise has generally been treated in the literature using recursive Bayesian estimators which rely on an approximation of the a *posteriori* density of the state variables. These Bayesian approximations are discussed in the next section.

2.3 Bayesian Approximations

Two problems are encountered when the the system is nonlinear or the a *priori* density is non-Gaussian. First, the integration in equation (2.4) is difficult

to carry out. Second, the moments are not easily obtained from equation (2.3). If the conditional density function cannot be computed analytically then the next best thing is to form accurate approximations of this density. Several numerical methods have been developed for approximation of the a *posteriori* density function. Some of these techniques are briefly discussed in this section.

Alspach and Sorenson [10,11] attempt to approximate the a *priori* density using a sum of Gaussian distributions. They apply their system to problems involving nonlinear state and measurement systems in white Gaussian noise. The procedure results in parallel operation of several Kalman filters. There are as many Kalman filters as there are terms in the Gaussian sum. The convex combination of these filters is formed to obtain the a *posteriori* density.

Sorenson and Stubberud [12] approximate the a *posteriori* density using an Edgeworth expansion. Using perturbation techniques the plant and measurement systems are described as quadratic equations with additive white Gaussian noise. Recursion relations are derived for a finite number of the moments of the Edgeworth expansions and these relations are assumed to describe the set of sufficient statistics for the system.

Bucy and Senne [13] use a crude convolution summation involving an ellipsoid tracking technique to determine the important points to include in the summation for the conditional density. They assume that the conditional densities of interest are sufficiently non-Gaussian so that a finite number of moments make for a poor representation of them. They store the densities as a vector of point masses relative to a rectangular grid which is free to be rotated and translated in the state space of the dynamical system.

Another method is to use spline filters [14,15] to construct the *a posteriori* density. Masi *et al.* [17] studies nonlinear discrete time filtering problems using the Bayesian approach. The solution to the filtering problem is given in terms of a generalized finite-dimensional filter in a sense that the generalized *a posteriori* conditional PDF is representable as a linear combination of distributions belonging to a given parameterized family, where the number of terms in the combination may possibly vary with time. Using this concept they are able to derive a technique to obtain exact recursive solutions for various linear models with non-Gaussian disturbances, as well as for one non-linear model with Gaussian disturbances.

All of these methods involve numerical approximations to the actual *a posteriori* density. The major limitation to these approaches is the computation time required for implementation.

2.4 Nonlinear Filtering in Gaussian Noise

This section presents a discussion of filtering methods for nonlinear systems that are described by the equation

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \Gamma_{k-1}\mathbf{w}_{k-1}, \tag{2.17}$$

with measurement model

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \tag{2.18}$$

where $\mathbf{v}_k$ and $\mathbf{w}_{k-1}$ are mutually independent white Gaussian noise sequences as described by equation (2.7).

In general, optimal Bayesian solutions cannot be expressed in closed form for this model, requiring methods for approximating optimal nonlinear filters. Several

nonlinear filters have been used for nonlinear systems in Gaussian noise. All of these filters are based on the model of the filtered state being a linear function of the innovations sequence as in equation (2.13). These suboptimal nonlinear filters include the extended Kalman filter, the modified second order Gaussian filter, the locally iterated Kalman filter, and the minimum variance filter. These filters are described in sections (2.4.1 – 2.4.4) respectively. Jazwinski [9] points out that it is difficult to assess a *priori* the effects of the approximations made by these nonlinear techniques, and their value in a particular problem must be determined by simulations.

2.4.1 Extended Kalman Filter (EKF)

The extended Kalman filter is obtained by making Gaussian assumptions about the a *posteriori* densities and by extending the plant and measurement nonlinearities in a Taylor series including first order terms.

The prediction error is defined as

$$\tilde{\mathbf{x}}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \tag{2.19}$$

and the filter error as

$$\tilde{\mathbf{x}}_{k-1|k-1} = \mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}. \tag{2.20}$$

If $\mathbf{f}_{k-1}$ is expanded about the current estimate, $\hat{\mathbf{x}}_{k-1|k-1}$, then the first order approximation is

$$\mathbf{f}_{k-1}(\mathbf{x}_{k-1}) \approx \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1}) + F_{k-1}\tilde{\mathbf{x}}_{k-1|k-1}, \tag{2.21}$$

where

$$F_{k-1} = \left.\frac{\partial \mathbf{f}_{k-1}(\mathbf{x}_{k-1})}{\partial \mathbf{x}_{k-1}}\right|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}}. \tag{2.22}$$

In this case $\hat{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1}) = \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1})$ and the prediction error becomes

$$\tilde{\mathbf{x}}_{k|k-1} = F_{k-1}\tilde{\mathbf{x}}_{k-1|k-1} + \Gamma_{k-1}\mathbf{w}_{k-1}. \tag{2.23}$$

This leads to the state prediction equations

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1}) \\ P_{k|k-1} &= E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}_{k|k-1}^T] \\ &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + \Gamma_{k-1}Q_{k-1}\Gamma_{k-1}^T. \end{aligned} \tag{2.24}$$

The measurement equation is linearized in a similar manner. The nonlinear function $\mathbf{h}_k(\mathbf{x}_k)$ is expanded about the predicted state $\hat{\mathbf{x}}_{k|k-1}$ to obtain

$$\mathbf{h}_k(\mathbf{x}_k) \approx \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) + H_k\tilde{\mathbf{x}}_{k|k-1} \tag{2.25}$$

where

$$H_k = \left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}. \tag{2.26}$$

The innovations vector is given by

$$\tilde{\mathbf{z}}_k = H_k\tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k. \tag{2.27}$$

The filter and gain equations have the same form as the linear Kalman filter and are given by

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + K_k\tilde{\mathbf{z}}_k \\ P_{k|k} &= (I_n - K_kH_k)P_{k|k-1} \\ K_k &= P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}. \end{aligned} \tag{2.28}$$

2.4.2 Gaussian Second Order (GSO) Filter

The Gaussian second order filter [6] is obtained by including the second order terms in the Taylor series expansion. In this filter it is assumed that all errors are Gaussian and therefore all odd moments are zero.

The expansions for $\mathbf{f}_{k-1}(.)$ and $\mathbf{h}_k(.)$ are given by

$$\begin{aligned}\mathbf{f}_{k-1}(\mathbf{x}_{k-1}) &\approx \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1}) + F_{k-1}\tilde{\mathbf{x}}_{k-1|k-1} + \frac{1}{2}\partial^2(\mathbf{f}_{k-1}, \tilde{\mathbf{x}}_{k-1|k-1}\tilde{\mathbf{x}}_{k-1|k-1}^T) \\ \mathbf{h}_k(\mathbf{x}_k) &\approx \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) + H_k\tilde{\mathbf{x}}_{k|k-1} + \frac{1}{2}\partial^2(\mathbf{h}_k, \tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}_{k|k-1}^T)\end{aligned} \tag{2.29}$$

where the operator $\partial^2(\mathbf{e}, B)$ for any function $\mathbf{e}(\mathbf{x})$ and any matrix B is a vector whose i^{th} element is defined by

$$\partial^2(\mathbf{e}, B) = \text{trace}\left\{\left[\frac{\partial^2 e_i}{\partial x_p \partial x_q}\right] B\right\}$$

for $1 \leq p \leq n$, $1 \leq q \leq n$. From (2.29) the estimates $\hat{\mathbf{f}}_{k-1}(\mathbf{x_{k-1}})$ and $\hat{\mathbf{h}}_k(\mathbf{x}_k)$ become

$$\begin{aligned}\hat{\mathbf{f}}_{k-1}(\mathbf{x_{k-1}}) &= \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1}) + \frac{1}{2}\,\partial^2(\mathbf{f}_{k-1}, P_{k-1|k-1})\Big|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}} \\ \hat{\mathbf{h}}_k(\mathbf{x}_k) &= \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) + \frac{1}{2}\,\partial^2(\mathbf{h}_k, P_{k|k-1})\Big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}.\end{aligned} \tag{2.30}$$

The innovations vector is now

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) - \frac{1}{2}\,\partial^2(\mathbf{h}_k, P_{k|k-1})\Big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}. \tag{2.31}$$

The GSO filter relations are given by

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1}) + \frac{1}{2}\,\partial^2(\mathbf{f}_{k-1}, P_{k-1|k-1})\Big|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}} \\
P_{k|k-1} &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + \Gamma_{k-1}Q_{k-1}\Gamma_{k-1}^T + A_{k-1} \\
\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + K_k\tilde{\mathbf{z}}_k \\
P_{k|k} &= (I_n - K_kH_k)P_{k|k-1} \\
K_k &= P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k + B_k)^{-1}.
\end{aligned} \tag{2.32}$$

In general, the matrices A_{k-1} and B_k contain fourth order moments. It is assumed that the prediction and filter PDF's are Gaussian for the development of the Gaussian second order filter. This assumption leads to the approximations

$$\begin{aligned}
A_{k-1_{ij}} &\approx \frac{1}{4}\left[\sum_{p,q,m,n}\frac{\partial^2 f_{k-1_j}}{\partial x_p\,\partial x_q}(c_{pm}c_{qn} + c_{pn}c_{qm})\frac{\partial^2 f_{k-1_i}}{\partial x_m\,\partial x_n}\right]_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}} \\
B_{k_{ij}} &\approx \frac{1}{4}\left[\sum_{p,q,m,n}\frac{\partial^2 h_{k_j}}{\partial x_p\,\partial x_q}(d_{pm}d_{qn} + d_{pn}d_{qm})\frac{\partial^2 h_{k_i}}{\partial x_m\,\partial x_n}\right]_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}
\end{aligned} \tag{2.33}$$

where f_{k-1_i} denotes the i^{th} element of $\mathbf{f}_{k-1}(.)$, h_{k_i} denotes the i^{th} element of $\mathbf{h}_k(.)$, the c's are elements of $P_{k|k}$, and the d's are elements of $P_{k|k-1}$.

Another approximation which was developed by Jazwinski [18] and Bass *et al.* [19] is the truncated second order filter. Third and higher order central moments are assumed to be zero in this filter. This results in slightly different equations than that shown above for the GSO filter. This filter is appropriate if the conditional density is almost symmetrical and concentrated near its mean. Still another version of second order filters is the modified Gaussian second order filter [20].

The appropriateness of the type of filter to use is dependent on the nonlinear-

ities in the system and can only be accurately determined by Monte Carlo simulations. It is difficult to analytically determine the effects of nonlinearities. However, Jazwinski [9] points out that, in general, nonlinear effects appear to be significant when noise inputs are small and the estimation error variance is large. Large noise inputs effectively mask nonlinearities. In addition he claims that measurement nonlinearities become significant whenever they are comparable to, or larger than, the measurement noise. Thus, if the measurement noise is small, neglected measurement nonlinearities tend to bias the estimate and result in incorrect weighting of the observations.

2.4.3 Locally Iterated Kalman Filter (LIKF)

The locally iterated Kalman filter is an enhanced version of the extended Kalman filter where, at each step of the iteration procedure, the measurement nonlinearity is linearized about the state estimate obtained from the EKF equations. This filter was first introduced by Denham and Pines [21]. The procedure is to repetitively calculate $\hat{\mathbf{x}}_{k|k}, K_k$, and $P_{k|k}$, each time linearizing about the most recent estimate. To develop this algorithm, denote the i^{th} estimate of $\mathbf{x}_{k|k}$ by $\hat{\mathbf{x}}_{k|k}(i)$ with $\hat{\mathbf{x}}_{k|k}(0) = \hat{\mathbf{x}}_{k|k-1}$ and expand $\mathbf{h}_k(\mathbf{x}_k)$ from equation (2.25) in the form

$$\mathbf{h}_k(\mathbf{x}_k) = \mathbf{h}_k(\hat{\mathbf{x}}_{k|k}(i)) + H_k\tilde{\mathbf{x}}_{k|k}(i)$$

where

$$H_k = \left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k}(i)}$$

$$\tilde{\mathbf{x}}_{k|k}(i) = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}(i).$$

The following recursion relations are developed [6]

$$\hat{\mathbf{x}}_{k|k}(i+1) = \hat{\mathbf{x}}_{k|k-1} + K_k(i)[\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k}(i)) - H_k(\hat{\mathbf{x}}_{k|k-1} - \hat{\mathbf{x}}_{k|k}(i))$$

$$P_{k|k}(i) = (I_n - K_k(i)H_k)P_{k|k-1} \tag{2.34}$$

$$K_k(i) = P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}$$

where $i = 0, 1, \cdots$. The number of repetitions of the calculations shown above can be determined by requiring the magnitude of the difference between successive state estimates to be less than some small number.

Jazwinski [9] gives the local iterated Kalman filter a probabilistic interpretation. Between observations, the conditional mean and covariance matrix propagate according to first order, nonlinear theory. At an observation, assuming the *a priori* density is Gaussian, the filter solves for the conditional mode of the posterior density. The conditional covariance matrix is then computed according to first order theory. The conditional mode is then used for the conditional mean.

Some disadvantages of the LIKF are pointed out by Andrade Netto *et al.* [22]:

(1) The iteration scheme may converge very slowly. This may occur where the initial guess $\mathbf{x}_{k|k-1}$ lies near extrema of the function $\mathbf{h}_k(.)$

(2) The *a posteriori* density may be multimodal and the iteration procedure may converge to local modes if it converges at all.

Another iteration scheme involves global iteration [9]. After processing the data $(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_k)$, starting with the initial values $\hat{\mathbf{x}}_0$, and P_0, the filtering operation is completed with estimates $\hat{\mathbf{x}}_{k|k}$ and $P_{k|k}$. Then, assuming $\mathbf{w}_k = 0$ the

backward filter is implemented with $\hat{\mathbf{x}}_{k|k}$ and $P_{k|k}$ as initial conditions. This gives smoothed estimates $\hat{\mathbf{x}}_{0|k}$ and $P_{0|k}$. The data is then processed with the forward filter again starting with $\hat{\mathbf{x}}_{0|k}$ and P_0. This has the effect of changing the initial statistic $E[\mathbf{x}_0]$.

2.4.4 Minimum Variance Filter (MVF)

The nonlinear filtering techniques discussed thus far are all based on a Taylor series expansion of the nonlinear equations about the most recent estimate. As such, these filters are subject to the inherent problems of local linearizations and may lead to poor performance. Liang and Christenson [23] developed filtering and smoothing algorithms which give exact estimates at each iteration of the filter. They have shown that for certain nonlinear functions such as polynomial nonlinearities, exponential functions, and sinusoids, exact expressions for the state estimates can be obtained and used in the filter relations in place of the usual approximations. At each step in the operation they assume that the prediction and filter errors are Gaussian. They have compared their filter to the EKF and other filters using numerical examples and claim that their filter performs much better than the EKF for large initial error variances.

The basic premise is that $E[\mathbf{f}_{k-1}(\mathbf{x}_{k-1})]$ and $E[\mathbf{h}_k(\mathbf{x}_k)]$ can be determined analytically such that

$$\hat{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1}) = E[\mathbf{f}_{k-1}(\mathbf{x}_{k-1})]_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}}$$

$$\hat{\mathbf{h}}_k(\mathbf{x}_k) = E[\mathbf{h}_k(\mathbf{x}_k)]_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}.$$

The innovations vector has the form

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \hat{\mathbf{h}}_k(\mathbf{x}_k).$$

The filter equations have the general form

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \hat{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1}) \\
P_{k|k-1} &= E[\tilde{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1})\tilde{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1})^T] + \Gamma_{k-1}Q_{k-1}\Gamma_{k-1}^T \\
\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + K_k\tilde{\mathbf{z}}_k \\
K_k &= E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{h}}_k(\mathbf{x}_k)^T]\,(R_k + E[\tilde{\mathbf{h}}_k(\mathbf{x}_k)\tilde{\mathbf{h}}_k(\mathbf{x}_k)^T])^T \\
P_{k|k} &= P_{k|k-1} - K_k\,E[\tilde{\mathbf{h}}_k(\mathbf{x}_k)\tilde{\mathbf{x}}_{k|k-1}^T]
\end{aligned} \tag{2.35}$$

where

$$\begin{aligned}
\tilde{\mathbf{x}}_{k|k-1} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \\
\tilde{\mathbf{h}}_k(\mathbf{x}_k) &= \mathbf{h}_k(\mathbf{x}_k) - \hat{\mathbf{h}}_k(\mathbf{x}_k) \\
\tilde{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1}) &= \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) - \hat{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1}).
\end{aligned} \tag{2.36}$$

Analytical expressions for $E[\mathbf{x}_k\mathbf{f}_k(\mathbf{x}_k)^T]$, $E[\mathbf{x}_k\mathbf{h}_k(\mathbf{x}_k)^T]$, $E[\mathbf{f}_k(\mathbf{x}_k)\mathbf{f}_k(\mathbf{x}_k)^T]$, and $E[\mathbf{h}_k(\mathbf{x}_k)\mathbf{h}_k(\mathbf{x}_k)^T]$ are required in order to form exact expressions for the filter equations.

It is important to note that the filter equations developed by Liang and Christenson [23] have been presented before (e.g. Jazwinski [9]). However, their contribution is the development of exact expressions for specific types of nonlinearities including polynomial, exponential, and sinusoid nonlinearities, assuming Gaussian *a posteriori* density functions. Liang [24] gives general expressions for the

probability density functions for these types of nonlinearities. Liang [25] evaluates several system models with the standard EKF and the MVF. He concludes that, in general, the MVF performs much better than the EKF for large initial error variances and small noise variances. However, when the initial variances are small, and the noise variances are not too small, the EKF can be expected to perform about as well as any other filter. He also claims that when the level of noise inputs is large enough to effectively cover the effects of nonlinearities, no particular filter can be said to be consistently superior to any other filter. In most cases, however, the MVF should outperform all other nonlinear filters considered.

Kramer and Sorenson [3] compare the performance of the MVF to the optimal Bayesian estimator for a specific bilinear model. They found that there is a wide margin between the performance of the suboptimal filter (MVF) and the optimal Bayesian filter. They generalize that when the level of noise inputs is large enough to mask the effects of nonlinearities, point estimators such as the MVF and EKF tend to perform close to the optimum. However, they may be quite sensitive to initial conditions. However, the MVF still fails to capture important features of the a *posteriori* densities.

2.5 Estimation in Non-Gaussian Noise

Most of the work done in filtering non-Gaussian Noise has been done from the Bayesian point of view. These techniques are discussed in sections 2.2 and 2.3. However, with a few notable exceptions there very little work has been done in the area of linear filtering of systems with non-Gaussian plant noise, measurement noise, or initial error variances. Some of the approximations found in the literature will be discussed in this section.

Masreliez [26] developed two methods for non-Gaussian filtering. One filter is used for situations in which the observation prediction density is approximately Gaussian at each stage, but the observations disturbances are non-Gaussian. He develops the filter using a nonlinear ("score") function of the innovations vector. The second filter applies to the systems with non-Gaussian plant noise but linear measurement noise. They compared their filter to the exact Bayesian minimum variance filter developed by Alspach and Sorenson [10] using a sum of two Gaussian distributions. Simulation runs indicated that the exact MV filter and the approximations presented in this paper coincide and that these filters outperform the Kalman filter. However, the author notes that the score function is very sensitive to small errors in the density approximations. They suggest that *ad hoc* type filters may be constructed to approximate the densities.

Another approach is taken by Rao and Yar [27]. In their paper they developed two filters for tracking nonlinear processes for scalar models. In the first technique, called the polynomial filter, they used a general n^{th} power of the innovations process for symmetrical plant noise, measurement noise, and initial variance distributions to develop relations that could be used to obtain the filter gain(s). However, they consider only the scalar cases with symmetrical distributions. The second filter, labeled the measurement noise dependent filter, is based on a general nonlinear model of the innovations. This filter is constrained by the fact that one must have exact knowledge of the measurement noise distribution.

Verriest [28] proposed a filter that would operate in multiplicative or non-Gaussian noise. This filter was set up for symmetric distributions, and equations were developed based on linear approximations. However, these equations were not verified with numerical results.

An exact formula for computing the conditional mean has been derived by Daum [29] for discrete time observations with non-Gaussian measurement noise. The derivation of this formula is based on a certain homotopy function. He mentions that in order to use his formula a conditional expectation must be computable, which is not generally the case in nonlinear estimation problems.

Chapter 3

High Order Filters for Estimation in Non-Gaussian Noise

In this chapter high order vector filter equations are developed for estimation in non-Gaussian noise. The difference between the filters developed here and the standard Kalman filter is that the filter equation contains nonlinear functions of the innovations process. These filters are general in that the initial state covariance, the measurement noise covariance, and the process noise covariance can all have non-Gaussian distributions. Two filter structures are developed. The first filter is designed for systems with asymmetric probability densities. The second is designed for systems with symmetric probability densities. Experimental evaluation of these filters for estimation in non-Gaussian noise, formed from Gaussian sum distributions, shows that these filters perform much better than the standard Kalman filter, and close to the optimal Bayesian estimator.

The new filters are referred to as high order filters (HOFs). For both of these filters it is assumed that the 5^{th} and higher order moments of all densities are negligible. As such, these filters are approximations of the optimal minimum variance solution. However, it is shown through simulation experiments that these filters can approach the performance of the optimal minimum variance filter under certain conditions. The performance of the HOFs is compared to the standard Kalman filter, which uses only first and second moments, and to the optimal Bayesian estimator. The Gaussian sum distributions, for which the optimal Bayesian estimator has been derived by Sorenson and Alspach [10], were used as the test-bed for comparison. Unlike the measurement noise dependent (MND) filter described in [26], which requires complete knowledge of the entire *a priori* densities, the HOFs re-

quire knowledge of only a finite number of moments of these densities. The optimal Bayesian estimator of Sorenson and Alspach [10] also requires the accurate knowledge of the *a priori* densities so that an approximation can be made using Gaussian sums. All techniques previously developed for non-Gaussian filtering are computationally intensive. The HOFs developed here share that characteristic. However, they are much less computationally intensive than the Gaussian sum filter.

3.1 System Model

Consider the problem of estimating the n-dimensional state vector $\mathbf{x}_k$ from K measurements of the m-dimensional vector $\mathbf{z}_k$. The linear plant and measurement equations have the form

$$\begin{aligned} \mathbf{x}_k &= \Phi_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{z}_k &= H_k\mathbf{x}_k + \mathbf{v}_k \end{aligned} \tag{3.1}$$

where $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are mutually independent, white, zero-mean, possibly non-Gaussian random sequences. The uncertainty in the initial estimate $\hat{\mathbf{x}}_0$ may also have a non-Gaussian distribution and is independent from $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$. It is also assumed that the 2^{nd} through 4^{th} moments of the distributions of $\hat{\mathbf{x}}_0$, $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are known.

The Kronecker product operator $\otimes$ [30] is implemented in order to use 2-dimension matrix operations throughout this derivation. The Kronecker product of

an $m \times n$ matrix A with a matrix B is defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \tag{3.2}$$

where a_{ij} is the ij^{th} element of the matrix A. The kronecker product has higher algebraic order than multiplication.

For arbitrary matrices A and B and arbitrary column vectors $\mathbf{a}$ and $\mathbf{b}$, the kronecker product has the following properties:

$$\begin{aligned} (A\mathbf{a}) \otimes (B\mathbf{b}) &= (A \otimes B)(\mathbf{a} \otimes \mathbf{b}) \\ (A\mathbf{a}) \otimes \mathbf{b} &= (A \otimes I_m)(\mathbf{a} \otimes \mathbf{b}) \\ (A \otimes B)^T &= A^T \otimes B^T \\ \mathbf{a} \otimes \mathbf{b}^T &= \mathbf{b}^T \otimes \mathbf{a} = \mathbf{a}\mathbf{b}^T \end{aligned} \tag{3.3}$$

where $\mathbf{b}$ is an m-dimensional vector, and I_m is the $m \times m$-dimensional identity matrix.

In the development to follow the column stack operator is also used. If an $n \times n$ matrix A consists of columns $\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_n}$ then the column stack of A is defined by

$$\text{cst}(A) \equiv [\mathbf{a_1^T a_2^T} \cdots \mathbf{a_n^T}]^T \tag{3.4}$$

$\text{cst}(A)$ is dimensioned $nn \times 1$. If $A = E[\mathbf{x}\mathbf{x^T}]$ then $\text{cst}(A) = E[\mathbf{x} \otimes \mathbf{x}]$, where $E[.]$ is the expectation operator.

The semi-column stack is defined as follows: if the matrix A is dimensioned $nn \times nn$, consisting of columns $\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_{nn}}$ then the semi-column stack of A is given by

$$\text{scst}(A) \equiv \begin{bmatrix} \mathbf{a_1} & \mathbf{a_{n+1}} & \cdots & \mathbf{a_{(n-1)*n+1}} \\ \mathbf{a_2} & \mathbf{a_{n+2}} & \cdots & \mathbf{a_{(n-1)*n+2}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a_n} & \mathbf{a_{n+n}} & \cdots & \mathbf{a_{(n-1)*n+n}} \end{bmatrix} \tag{3.5}$$

$\text{scst}(A)$ is dimensioned $nnn \times n$. If $A = E[\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}^\mathbf{T} \otimes \mathbf{x}^\mathbf{T}]$ then $\text{scst}(A) = E[\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}^\mathbf{T}]$.

The $2^{nd}, 3^{rd}$, and 4^{th} moments of the random vector $\mathbf{w}_k$ are given as

$$E[\mathbf{w}_k \otimes \mathbf{w}_j^T] = Q_k^{(2)} \delta_{kj}$$

$$E[\mathbf{w}_k \otimes \mathbf{w}_j^T \otimes \mathbf{w}_l] = Q_k^{(3)} \delta_{kjl}$$

$$E[\mathbf{w}_k \otimes \mathbf{w}_j^T \otimes \mathbf{w}_l \otimes \mathbf{w}_m^T] = Q_k^{(4)} \delta_{kjlm}$$

Similarly, the $2^{nd}, 3^{rd}$, and 4^{th} moments of the random vector $\mathbf{v}_k$ are given by $R_k^{(2)} \delta_{kj}$, $R_k^{(3)} \delta_{kjl}$, and $R_k^{(4)} \delta_{kjlm}$. The moments of the initial estimation error are given by $P_0^{(2)}$, $P_0^{(3)}$, and $P_0^{(4)}$.

Let the prediction error $\tilde{\mathbf{x}}_{k|k-1}$ and the filtered error $\tilde{\mathbf{x}}_{k|k}$ be defined as

$$\begin{aligned} \tilde{\mathbf{x}}_{k|k-1} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \\ \tilde{\mathbf{x}}_{k|k} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k} \end{aligned} \tag{3.6}$$

where the hat indicates expected value. The innovations vector is given by

$$\begin{aligned} \tilde{\mathbf{z}}_k &= \mathbf{z}_k - H_k \hat{\mathbf{x}}_{k|k-1} \\ &= H_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k \end{aligned} \tag{3.7}$$

It can be shown [27] that $E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})|\mathbf{z}_k]$ is a function of only $\tilde{\mathbf{z}}_k$ so that

$$E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})|\mathbf{z}_k] = \sum_{i=0}^{\infty} K_k^{(i)} \tilde{\mathbf{z}}_k^{\otimes i}$$

where the superscript $\otimes i$ denotes the i^{th} kronecker product of the vector $\tilde{\mathbf{z}}_k$. $K_k^{(i)}$ denotes the i^{th} order filter gain, which has dimension $n \times m^i$. It follows that

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \sum_{i=1}^{\infty} K_k^{(i)} \tilde{\mathbf{z}}_k^{\otimes i}$$

Using (3.6) the expression for the filter error becomes

$$\tilde{\mathbf{x}}_{k|k} = \tilde{\mathbf{x}}_{k|k-1} - \sum_{i=1}^{\infty} K_k^{(i)} \tilde{\mathbf{z}}_k^{\otimes i}. \tag{3.8}$$

By setting $K_k^{(i)} = 0$, for $i > 1$, the standard linear Kalman filter results. In order to bound the equations for the derivation of the high order filters it is assumed that $\tilde{\mathbf{z}}_k^{\otimes i}$ is negligible for $i > I$. The truncated relation now becomes

$$\tilde{\mathbf{x}}_{k|k} = \tilde{\mathbf{x}}_{k|k-1} - \sum_{i=0}^{I} K_k^{(i)} \tilde{\mathbf{z}}_k^{\otimes i}. \tag{3.9}$$

Equation (3.9) forms the basis for the development of the HOFs.

3.2 Non-Gaussian Filtering for Asymmetrical Distributions

The non-Gaussian filter for asymmetrical distributions is derived by letting $I = 2$ in (3.9) and obtain the filter error

$$\tilde{\mathbf{x}}_{k|k} = \tilde{\mathbf{x}}_{k|k-1} - K_k^{(0)} - K_k^{(1)} \tilde{\mathbf{z}}_k - K_k^{(2)} \tilde{\mathbf{z}}_k^{\otimes 2} \tag{3.10}$$

It is required that $E[\tilde{\mathbf{x}}_{k|k-1}] = E[\mathbf{v}_k] = 0$, since the estimator must be unbiased,

and using (3.7) in (3.10)

$$E[\tilde{\mathbf{x}}_{k|k-1}] = 0 = -K_k^{(0)} - K_k^{(2)} E[\tilde{\mathbf{z}}_k^{\otimes 2}] \tag{3.11}$$

where $E[\tilde{\mathbf{z}}_k^{\otimes 2}] = \mathrm{cst}(H_k P_{k|k-1}^{(2)} H_k^T + R_k^{(2)})$. Substituting (3.11) into (3.10) yields

$$\tilde{\mathbf{x}}_{k|k} = \tilde{\mathbf{x}}_{k|k-1} - K_k^{(1)} \tilde{\mathbf{z}}_k - K_k^{(2)} \tilde{\zeta}_k \tag{3.12}$$

where $\tilde{\zeta}_k$ is defined for notational convenience as

$$\tilde{\zeta}_k = (\tilde{\mathbf{z}}_k^{\otimes 2} - E[\tilde{\mathbf{z}}_k^{\otimes 2}]) = ((\tilde{\mathbf{z}}_k \otimes \tilde{\mathbf{z}}_k) - \mathrm{cst}(H_k P_{k|k-1}^{(2)} H_k^T + R_k^{(2)})) \tag{3.13}$$

which is a second order function of the innovations with $E[\tilde{\zeta}_k] = 0$. The corresponding filter equation is

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k^{(1)} \tilde{\mathbf{z}}_k + K_k^{(2)} \tilde{\zeta}_k \tag{3.14}$$

The formulas for the gains $K_k^{(1)}$ and $K_k^{(2)}$ result from the requirement for a minimum variance solution. Using (3.12) the equation for the variance of the a *posteriori* density becomes

$$\begin{aligned} P_{k|k}^{(2)} &= E[\tilde{\mathbf{x}}_{k|k} \tilde{\mathbf{x}}_{k|k}^T] \\ &= E[\tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^T] - E[\tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{z}}_k^T] K_k^{(1)\,T} - E[\tilde{\mathbf{x}}_{k|k-1} \tilde{\zeta}_k^T] K_k^{(2)\,T} \\ &\quad - K_k^{(1)} E[\tilde{\mathbf{z}}_k \tilde{\mathbf{x}}_{k|k-1}^T] + K_k^{(1)} E[\tilde{\mathbf{z}}_k \tilde{\mathbf{z}}_k^T] K_k^{(1)\,T} + K_k^{(1)} E[\tilde{\mathbf{z}}_k \tilde{\zeta}_k^T] K_k^{(2)\,T} \\ &\quad - K_k^{(2)} E[\tilde{\zeta}_k \tilde{\mathbf{x}}_{k|k-1}^T] + K_k^{(2)} E[\tilde{\zeta}_k \tilde{\mathbf{z}}_k^T] K_k^{(1)\,T} + K_k^{(2)} E[\tilde{\zeta}_k \tilde{\zeta}_k^T] K_k^{(2)\,T}. \end{aligned} \tag{3.15}$$

The gains are then determined from the matrix minimum principal [23] by evaluating

$$\frac{\partial \mathrm{trace}\{P_{k|k}^{(2)}\}}{\partial K_k^{(1)}} = 0, \quad \frac{\partial \mathrm{trace}\{P_{k|k}^{(2)}\}}{\partial K_k^{(2)}} = 0. \tag{3.16}$$

Carrying out these operations on (3.15) yields

$$\begin{aligned} K_k^{(1)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T] - K_k^{(2)}E[\tilde{\zeta}_k\tilde{\mathbf{z}}_k^T])E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T]^{-1} \\ K_k^{(2)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\zeta}_k^T] - K_k^{(1)}E[\tilde{\mathbf{z}}_k\tilde{\zeta}_k^T])E[\tilde{\zeta}_k\tilde{\zeta}_k^T]^{-1}. \end{aligned} \tag{3.17}$$

It is observed that $E[\tilde{\zeta}_k\tilde{\zeta}_k^T]$ is singular. This is a consequence of the fact that $\tilde{\zeta}_k$ contains repeated terms. For example, if the dimensionality M of the innovations vector $\tilde{\zeta}_k$ is 2, then the term $(\tilde{z}_k(1)\tilde{z}_k(2) - E[\tilde{z}_k(1)\tilde{z}_k(2)])$, where $\tilde{z}_k(j)$ is the j^{th} of $\tilde{\mathbf{z}}_k$, appears twice in $\tilde{\zeta}_k$. The number of repeated terms is a function of the dimensionality M of the innovations vector. To avoid this singularity, define the collapsed vector $\tilde{\zeta}_{kc}$ such that

$$\tilde{\zeta}_{kc} \equiv T_M\tilde{\zeta}_k \tag{3.18}$$

where T_M is a matrix of 1's and 0's designed to eliminate redundant columns or rows from $\tilde{\zeta}_k$. For example if $M = 2$

$$T_M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.19}$$

Since $\tilde{\zeta}_{kc}$ does not contain repeated terms, $[\tilde{\zeta}_{kc}\tilde{\zeta}_{kc}^T]$ is nonsingular. Let $K_{kc}^{(2)}$ denote the collapsed gain associated with replacing $\tilde{\zeta}_k$ with $\tilde{\zeta}_{kc}$ in (3.17). Solving for $K_k^{(1)}$ and $K_{kc}^{(2)}$ yields

$$\begin{aligned} K_k^{(1)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T] - E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\zeta}_{kc}^T]\,E[\tilde{\zeta}_{kc}\tilde{\zeta}_{kc}^T]^{-1}\,E[\tilde{\zeta}_{kc}\tilde{\mathbf{z}}_k^T]) \\ &\quad \times (E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T] - E[\tilde{\mathbf{z}}_k\tilde{\zeta}_{kc}^T]\,E[\tilde{\zeta}_{kc}\tilde{\zeta}_{kc}^T]^{-1}\,E[\tilde{\zeta}_{kc}\tilde{\mathbf{z}}_k^T])^{-1} \\ K_{kc}^{(2)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\zeta}_{kc}^T] - E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T]\,E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T]^{-1}\,E[\tilde{\mathbf{z}}_k\tilde{\zeta}_{kc}^T]) \\ &\quad \times (E[\tilde{\zeta}_{kc}\tilde{\zeta}_{kc}^T] - E[\tilde{\zeta}_{kc}\tilde{\mathbf{z}}_k^T]E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T]^{-1}\,E[\tilde{\mathbf{z}}_k\tilde{\zeta}_{kc}^T])^{-1}. \end{aligned} \tag{3.20}$$

Equation (3.14) requires that

$$K_k^{(2)}\tilde{\zeta}_k = K_{kc}^{(2)}\tilde{\zeta}_{kc} \tag{3.21}$$

Using (3.18) in (3.21), $K_k^{(2)}$ is then obtained from

$$K_k^{(2)} = K_{kc}^{(2)} T_M \tag{3.22}$$

where

$$E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T] = P_{k|k-1}^{(2)}\, H_k^T \tag{3.23}$$

$$E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\zeta}_k^T] = P_{k|k-1}^{(3)\,T}\, H_k^T \otimes H_k^T \tag{3.24}$$

$$E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T] = H_k\, P_{k|k-1}^{(2)}\, H_k^T + R_k^{(2)} \tag{3.25}$$

$$E[\tilde{\mathbf{z}}_k\tilde{\zeta}_k^T] = E[\tilde{\zeta}_k\tilde{\mathbf{z}}_k^T]^T = H_k\, P_{k|k-1}^{(3)\,T}\, H_k^T \otimes H_k^T + R_k^{(3)\,T} \tag{3.26}$$

$$\begin{aligned} E[\tilde{\zeta}_k\tilde{\zeta}_k^T] &= H_k \otimes H_k \,\{P_{k|k-1}^{(4)} - \mathrm{cst}\,(P_{k|k-1}^{(2)})\mathrm{cst}(P_{k|k-1}^{(2)})^T\}\, H_k^T \otimes H_k^T \\ &+ R_k^{(4)} - \mathrm{cst}(R_k^{(2)})\,\mathrm{cst}(R_k^{(2)})^T \\ &+ H_k \otimes I_m\, P_{k|k-1}^{(2)} \otimes R_k^{(2)}\, H_k^T \otimes I_m \\ &+ H_k \otimes I_m\, E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}_k^T \otimes \tilde{\mathbf{x}}_{k|k-1}^T]\, I_m \otimes H_k^T \\ &+ I_m \otimes H_k\, R_k^{(2)} \otimes P_{k|k-1}^{(2)}\, I_m \otimes H_k^T \\ &+ I_m \otimes H_k\, E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1}^T \otimes \mathbf{v}_k^T]\, H_k^T \otimes I_m \end{aligned} \tag{3.27}$$

It can easily be shown that if all 3^{rd} moments are zero then $K_k^{(2)} = 0$ and $K_k^{(1)}$ reduces to the gain for the standard Kalman filter.

Using the state model (3.1) the prediction equation becomes

$$\hat{\mathbf{x}}_{k|k-1} = \Phi_{k-1}\hat{\mathbf{x}}_{k-1|k-1}. \tag{3.28}$$

The corresponding prediction error from (3.6) is given by

$$\tilde{\mathbf{x}}_{k|k-1} = \Phi_{k-1}\tilde{\mathbf{x}}_{k-1|k-1} + \mathbf{w}_{k-1}. \tag{3.29}$$

The prediction moments are then be evaluated as

$$\begin{aligned} P^{(2)}_{k|k-1} &= E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}^T_{k|k-1}] \\ &= \Phi_{k-1} P^{(2)}_{k-1|k-1} \Phi^T_{k-1} + Q^{(2)}_{k-1} \end{aligned} \tag{3.30}$$

$$\begin{aligned} P^{(3)}_{k|k-1} &= E[(\tilde{\mathbf{x}}^{\otimes 2}_{k|k-1})\tilde{\mathbf{x}}^T_{k|k-1}] \\ &= \Phi_{k-1} \otimes \Phi_{k-1}\; P^{(3)}_{k-1|k-1}\; \Phi^T_{k-1} \;+\; Q^{(3)}_{k-1} \end{aligned} \tag{3.31}$$

$$\begin{aligned} P^{(4)}_{k|k-1} &= E[(\tilde{\mathbf{x}}^{\otimes 2}_{k|k-1})\,(\tilde{\mathbf{x}}^{\otimes 2}_{k|k-1})^T] \\ &= \Phi_{k-1} \otimes \Phi_{k-1}\; P^{(4)}_{k-1|k-1}\; \Phi^T_{k-1} \otimes \Phi^T_{k-1} \;+\; Q^{(4)}_{k-1} \\ &+\; \Phi_{k-1} \otimes I_n\; P^{(2)}_{k-1|k-1} \otimes Q^{(2)}_{k-1}\; \Phi^T_{k-1} \otimes I_n \\ &+\; \Phi_{k-1} \otimes \Phi_{k-1}\; \mathrm{cst}(P^{(2)}_{k-1|k-1})\; \mathrm{cst}(Q^{(2)}_{k-1})^T \\ &+\; \Phi_{k-1} \otimes I_n\; E[\tilde{\mathbf{x}}_{k-1|k-1} \otimes \mathbf{w}^T_{k-1} \otimes \mathbf{w}_{k-1} \otimes \tilde{\mathbf{x}}_{k-1|k-1}]\; I_n \otimes \Phi^T_{k-1} \\ &+\; I_n \otimes \Phi_{k-1}\; E[\mathbf{w}_{k-1} \otimes \tilde{\mathbf{x}}^T_{k-1|k-1} \otimes \tilde{\mathbf{x}}_{k-1|k-1} \otimes \mathbf{w}_{k-1}]\; \Phi^T_{k-1} \otimes I_n \\ &+\; \mathrm{cst}(Q^{(2)}_{k-1})\; \mathrm{cst}(P^{(2)}_{k-1|k-1})^T\; \Phi^T_{k-1} \otimes \Phi^T_{k-1} \\ &+\; I_n \otimes \Phi_{k-1}\; Q^{(2)}_{k-1} \otimes P^{(2)}_{k-1|k-1}\; I_n \otimes \Phi^T_{k-1} \end{aligned} \tag{3.32}$$

where I_n is an n-dimensional identity matrix. Similarly, the moments of the filter error can be evaluated using equation (3.12). Let

$$A_k \equiv (I_n - K^{(1)}_k H_k) \tag{3.33}$$

The filter variance becomes

$$
\begin{aligned}
P^{(2)}_{k|k} &= E[\tilde{\mathbf{x}}_{k|k}\tilde{\mathbf{x}}^T_{k|k}] \\
&= A_k\, P^{(2)}_{k|k-1}\, A^T_k \;+\; K^{(1)}_k R^{(2)}_k K^{(1)\,T}_k \\
&- \; K^{(2)}_k\, H_k \otimes H_k\, P^{(3)}_{k|k-1}\, A^T_k - \; A_k {P^{(3)}_{k|k-1}}^T\, H^T_k \otimes H^T_k\, K^{(2)\,T}_k \\
&+ \; K^{(2)}_k\, R^{(3)}_k\, K^{(1)\,T}_k + \; K^{(1)}_k\, R^{(3)\,T}_k\, K^{(2)\,T}_k \\
&+ \; K^{(2)}_k\, H_k \otimes H_k\, \{P^{(4)}_{k|k-1} - \mathrm{cst}(P^{(2)}_{k|k-1})\mathrm{cst}(P^{(2)}_{k|k-1})^T\}\, H^T_k \otimes H^T_k\, K^{(2)\,T}_k \\
&+ \; K^{(2)}_k\, \{R^{(4)}_k - \mathrm{cst}(R^{(2)}_k)\mathrm{cst}(R^{(2)}_k)^T\}\, K^{(2)\,T}_k \\
&+ \; K^{(2)}_k\, H_k \otimes I_m\, P^{(2)}_{k|k-1} \otimes R^{(2)}_k\, H^T_k \otimes I_m\, K^{(2)\,T}_k \\
&+ \; K^{(2)}_k\, H_k \otimes I_m\, E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1}]\, I_m \otimes H^T_k\, K^{(2)\,T}_k \\
&+ \; K^{(2)}_k\, I_m \otimes H_k\, R^{(2)}_k \otimes P^{(2)}_{k|k-1}\, I_m \otimes H^T_k\, K^{(2)\,T}_k \\
&+ \; K^{(2)}_k\, I_m \otimes H_k\, E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k]\, H^T_k \otimes I_m\, K^{(2)\,T}_k
\end{aligned}
\tag{3.34}
$$

It is observed that the equation for the n^{th} filter error moment requires the availability of prediction and measurement error moments of order $2n$. This is a consequence of the fact that the filter error given by (3.12) is a second order function of the innovations. Since only the prediction order moments up to 4^{th} order are propagated, the equations for the 3^{rd} and 4^{th} order filter moments are truncated so that they contain only 3^{rd} and 4^{th} order functions of the prediction and measurement error moments. An alternative would be to completely expand the 3^{rd} and 4^{th} order filter moments in terms of all $2n$ prediction and measurement error moments and approximate the higher moments using suitable functions of the 2^{nd} through 4^{th} moments. The vector expansion becomes very unwieldy and is not included here. However, Section 3.5 contains the scalar expansions. Simulation experiments presented in Section 3.7 compare the truncated models to the nontruncated models which use higher order moment approximation.

With this restriction the 3^{rd} order filter moment becomes

$$
\begin{aligned}
P^{(3)}_{k|k} &= E[(\tilde{\mathbf{x}}^{\otimes 2}_{k|k})\tilde{\mathbf{x}}^T_{k|k}] \\
&= A_k \otimes A_k \; P^{(3)}_{k|k-1} \otimes A_k^T \; - K^{(1)}_k \otimes K^{(1)}_k \; R^{(3)}_k K^{(1)\,T}_k \\
&- A_k \otimes (K^{(2)}_k \; H_k \otimes H_k) \; \{\mathrm{scst}(P^{(4)}_{k|k-1}) \; - \; P^{(2)}_{k|k-1} \otimes \mathrm{cst}(P^{(2)}_{k|k-1})\} \; A_k^T \\
&- A_k \otimes A_k \; \{P^{(4)}_{k|k-1} \; - \; \mathrm{cst}(P^{(2)}_{k|k-1})\mathrm{cst}(P^{(2)}_{k|k-1})^T\} \; H_k^T \otimes H_k^T \; K^{(2)\,T}_k \\
&- (K^{(2)}_k \; H_k \otimes H_k) \otimes A_k \; \{\mathrm{scst}(P^{(4)}_{k|k-1}) \; - \; \mathrm{cst}(P^{(2)}_{k|k-1}) \otimes P^{(2)}_{k|k-1}\} \; A_k^T \\
&- K^{(1)}_k \otimes K^{(2)}_k \{\mathrm{scst}(R^{(4)}_k) \; - \; R^{(2)}_k \otimes \mathrm{cst}(R^{(2)}_k)\} \; K^{(1)\,T}_k \\
&- K^{(1)}_k \otimes K^{(1)}_k \; \{R^{(4)}_k \; - \; \mathrm{cst}(R^{(2)}_k)\mathrm{cst}(R^{(2)}_k)^T\} \; K^{(2)\,T}_k \\
&- K^{(2)}_k \otimes K^{(1)}_k \{\mathrm{scst}(R^{(4)}_k) \; - \; \mathrm{cst}(R^{(2)}_k) \otimes R^{(2)}_k\} \; K^{(1)\,T}_k \\
&+ A_k \otimes (K^{(2)}_k \; H_k \otimes I_m) \; \mathrm{cst}(P^{(2)}_{k|k-1}) \otimes R^{(2)}_k \; K^{(1)\,T}_k \\
&+ A_k \otimes (K^{(2)}_k \; I_m \otimes H_k) \;\; E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}^T_k \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1}] \; K^{(1)\,T}_k \\
&+ A_k \otimes K^{(1)}_k \; P^{(2)}_{k|k-1} \otimes R^{(2)}_k \; H_k^T \otimes I_m \; K^{(2)\,T}_k \\
&+ A_k \otimes K^{(1)}_k \;\; E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}_k] \; I_m \otimes H_k^T \; K^{(2)\,T}_k \\
&+ K^{(1)}_k \otimes (K^{(2)}_k \; H_k \otimes I_m) \;\; E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k] \; A_k^T \\
&+ K^{(1)}_k \otimes (K^{(2)}_k \; I_m \otimes H_k) \;\; \mathrm{cst}(R^{(2)}_k) \otimes P^{(2)}_{k|k-1} \; A_k^T \\
&+ K^{(1)}_k \otimes A_k \; E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}_{k|k-1}] \; H_k^T \otimes I_m \; K^{(2)\,T}_k \\
&+ K^{(1)}_k \otimes A_k \; R^{(2)}_k \otimes P^{(2)}_{k|k-1} \; I_m \otimes H_k^T \; K^{(2)\,T}_k \\
&+ (K^{(2)}_k \; H_k \otimes I_m) \otimes K^{(1)}_k \; P^{(2)}_{k|k-1} \otimes \mathrm{cst}(R^{(2)}_k) \; A_k^T \\
&+ (K^{(2)}_k \; I_m \otimes H_k) \otimes K^{(1)}_k \;\; E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}_k] \; A_k^T \\
&+ (K^{(2)}_k \; H_k \otimes I_m) \otimes A_k \;\; E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}_{k|k-1}] \; K^{(1)\,T}_k \\
&+ (K^{(2)}_k \; I_m \otimes H_k) \otimes A_k \; R^{(2)}_k \otimes \mathrm{cst}(P^{(2)}_{k|k-1}) \; K^{(1)\,T}_k
\end{aligned}
\tag{3.35}
$$

The fourth order moment expansion is truncated so that it includes only

functions of 4^{th} order prediction and measurement error moments.

$$
\begin{aligned}
P_{k|k}^{(4)} &= E[(\tilde{\mathbf{x}}_{k|k}^{\otimes 2})\,(\tilde{\mathbf{x}}_{k|k}^{\otimes 2})^T] \\
&= A_k \otimes A_k\; P_{k|k-1}^{(4)}\; A_k^T \otimes A_k^T \\
&+\; A_k \otimes K_k^{(1)}\; P_{k|k-1}^{(2)} \otimes R_k^{(2)}\; A_k^T \otimes K_k^{(1)\,T} \\
&+\; A_k \otimes A_k\; \mathrm{cst}(P_{k|k-1}^{(2)})\; \mathrm{cst}(R_k^{(2)})^T\; K_k^{(1)\,T} \otimes K_k^{(1)\,T} \\
&+\; A_k \otimes K_k^{(1)}\; E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k^T \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1}^T]\; K_k^{(1)\,T} \otimes A_k^T \\
&+\; K_k^{(1)} \otimes A_k\; E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1}^T \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k^T]\; A_k^T \otimes K_k^{(1)\,T} \\
&+\; K_k^{(1)} \otimes K_k^{(1)}\; \mathrm{cst}(R_k^{(2)})\; \mathrm{cst}(P_{k|k-1}^{(2)})^T\; A_k^T \otimes A_k^T \\
&+\; K_k^{(1)} \otimes A_k\; R_k^{(2)} \otimes P_{k|k-1}^{(2)}\; K_k^{(1)\,T} \otimes A_k^T \\
&+\; K_k^{(1)} \otimes K_k^{(1)}\; R_k^{(4)}\; K_k^{(1)\,T} \otimes K_k^{(1)\,T}
\end{aligned}
\tag{3.36}
$$

The filter equations (3.14, 3.34-3.36), gain equations (3.20), and prediction equations (3.28, 3.30-3.32) constitute the discrete filter relations for non-Gaussian noise with arbitrary asymmetrical distributions. These relations are suboptimal in that they do not completely characterize the noise distributions since they use only the first four moments of the distributions.

3.3 Non-Gaussian Filtering for Symmetrical Distributions

The derivation for the non-Gaussian filter for symmetrical distributions follows the same general procedure as in the previous section. If the errors are assumed to have only even moments, then it can be shown that $K_k^{(i)} = 0$ for $i = 0, 2, 4, \cdots$ [5]. The truncated non-Gaussian filter for symmetrical distributions is obtained by letting $I = 3$ in equation (3.9) and obtain the filter error

$$
\tilde{\mathbf{x}}_{k|k} = \tilde{\mathbf{x}}_{k|k-1} - K_k^{(1)}\tilde{\mathbf{z}}_k - K_k^{(3)}(\tilde{\mathbf{z}}_k^{\otimes 3}). \tag{3.37}
$$

with corresponding filter equation

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k^{(1)}\tilde{\mathbf{z}}_k + K_k^{(3)}\,\tilde{\mathbf{z}}_k^{\otimes 3}. \tag{3.38}$$

The estimator is required to be unbiased. By definition all odd moments of the innovations are zero. Since $E[\tilde{\mathbf{x}}_{k|k-1}] = 0$, the expected value of the estimation error given in (3.37) is zero. For notational convenience let

$$\tilde{\alpha}_k = \tilde{\mathbf{z}}_k^{\otimes 3} \tag{3.39}$$

with $E[\tilde{\alpha}_k] = 0$.

The formulas for the gains $K_k^{(1)}$ and $K_k^{(3)}$ result from the requirement for a minimum variance solution. The variance of the *a posteriori* density function is given by

$$\begin{aligned} P_{k|k}^{(2)} &= E[\tilde{\mathbf{x}}_{k|k}\tilde{\mathbf{x}}_{k|k}^T] \\ &= E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}_{k|k-1}^T] - E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T]K_k^{(1)\,T} - E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\alpha}_k^T]K_k^{(3)\,T} \\ &- K_k^{(1)}E[\tilde{\mathbf{z}}_k\tilde{\mathbf{x}}_{k|k-1}^T] + K_k^{(1)}E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T]K_k^{(1)\,T} + K_k^{(1)}E[\tilde{\mathbf{z}}_k\tilde{\alpha}_k^T]K_k^{(3)\,T} \\ &- K_k^{(3)}E[\tilde{\alpha}_k\tilde{\mathbf{x}}_{k|k-1}^T] + K_k^{(3)}E[\tilde{\alpha}_k\tilde{\mathbf{z}}_k^T]K_k^{(1)\,T} + K_k^{(3)}E[\tilde{\alpha}_k\tilde{\alpha}_k^T]K_k^{(3)\,T}. \end{aligned} \tag{3.40}$$

From the matrix minimum principal [23]

$$\frac{\partial \text{trace}\,\{P_{k|k}^{(2)}\}}{\partial K_k^{(1)}} = 0, \quad \frac{\partial \text{trace}\,\{P_{k|k}^{(2)}\}}{\partial K_k^{(3)}} = 0. \tag{3.41}$$

Carrying out these operations on (3.40)

$$\begin{aligned} K_k^{(1)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T] - K_k^{(3)}E[\tilde{\alpha}_k\tilde{\mathbf{z}}_k^T]) \times E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T]^{-1} \\ K_k^{(3)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\alpha}_k^T] - K_k^{(1)}E[\tilde{\mathbf{z}}_k\tilde{\alpha}_k^T]) \times E[\tilde{\alpha}_k\tilde{\alpha}_k^T]^{-1}. \end{aligned} \tag{3.42}$$

Similar to the $E[\tilde{\zeta}_k\tilde{\zeta}_k^T]$ for the asymmetric filter, it is observed that $E[\tilde{\alpha}_k\tilde{\alpha}_k^T]$ is singular. A collapsed vector $\tilde{\alpha}_{k_c}$ is defined such that

$$\tilde{\alpha}_{k_c} \equiv U_M \tilde{\alpha}_k \tag{3.43}$$

where U_M is a matrix of 1's and 0's designed to extract only one of each term from $\tilde{\alpha}_k$. Since $\tilde{\alpha}_{k_c}$ does not contain repeated terms, $[\tilde{\alpha}_{k_c}\tilde{\alpha}_{k_c}^T]$ is nonsingular. Let $K_{k_c}^{(3)}$ denote the collapsed gain associated with replacing $\tilde{\alpha}_k$ with $\tilde{\alpha}_{k_c}$ in (3.42). Solving for $K_k^{(1)}$ and $K_k^{(3)}$ yields

$$\begin{aligned} K_k^{(1)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T] - E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\alpha}_{k_c}^T]\, E[\tilde{\alpha}_{k_c}\tilde{\alpha}_{k_c}^T]^{-1}\, E[\tilde{\alpha}_{k_c}\tilde{\mathbf{z}}_k^T]) \\ &\quad \times (E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T] - E[\tilde{\mathbf{z}}_k\tilde{\alpha}_{k_c}^T]\, E[\tilde{\alpha}_{k_c}\tilde{\alpha}_{k_c}^T]^{-1}\, E[\tilde{\alpha}_{k_c}\tilde{\mathbf{z}}_k^T])^{-1} \\ K_{k_c}^{(3)} &= (E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\alpha}_{k_c}^T] - E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T]\, E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T]^{-1}\, E[\tilde{\mathbf{z}}_k\tilde{\alpha}_{k_c}^T]) \\ &\quad \times (E[\tilde{\alpha}_{k_c}\tilde{\alpha}_{k_c}^T] - E[\tilde{\alpha}_{k_c}\tilde{\mathbf{z}}_k^T]E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T]^{-1}\, E[\tilde{\mathbf{z}}_k\tilde{\alpha}_{k_c}^T])^{-1}. \end{aligned} \tag{3.44}$$

Equation (3.38) requires that

$$K_k^{(3)}\tilde{\alpha}_k = K_{k_c}^{(3)}\tilde{\alpha}_{k_c} \tag{3.45}$$

Using (3.43) in (3.45), $K_k^{(3)}$ is then obtained from

$$K_k^{(3)} = K_{k_c}^{(3)}\, U_M. \tag{3.46}$$

Define the parameters

$$\begin{aligned} B_{k_1} &= H_k \otimes H_k \otimes H_k \\ B_{k_2} &= H_k \otimes H_k \otimes I_m \\ B_{k_3} &= H_k \otimes I_m \otimes H_k \\ B_{k_4} &= H_k \otimes I_m \otimes I_m \\ B_{k_5} &= I_m \otimes H_k \otimes H_k \\ B_{k_6} &= I_m \otimes H_k \otimes I_m \\ B_{k_7} &= I_m \otimes I_m \otimes H_k \\ B_{k_8} &= I_m \otimes I_m \otimes I_m \end{aligned} \tag{3.47}$$

The expectations given in equation (3.44) become

$$E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}_k^T] = P_{k|k-1}^{(2)}\, H_k^T \tag{3.48}$$

$$\begin{aligned} E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\alpha}_k^T] &= \mathrm{scst}(P_{k|k-1}^{(4)})^T\, B_{k_1}^T \;+\; P_{k|k-1}^{(2)} \otimes \mathrm{cst}(R_k^{(2)})^T\, B_{k_4}^T \\ &\quad + E[\mathbf{v}_k^T \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1}^T \otimes \mathbf{v}_k^T]\, B_{k_6}^T \;+\; \mathrm{cst}(R_k^{(2)})^T \otimes P_{k|k-1}^{(2)}\, B_{k_7}^T \end{aligned} \tag{3.49}$$

$$E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T] = H_k\, P_{k|k-1}^{(2)}\, H_k^T \;+\; R_k^{(2)} \tag{3.50}$$

$$\begin{aligned} E[\tilde{\mathbf{z}}_k\tilde{\alpha}_k^T] &= E[\tilde{\alpha}_k\tilde{\mathbf{z}}_k^T]^T \\ &= H_k\, \mathrm{scst}(P_{k|k-1}^{(4)})^T\, B_{k_1}^T \;+\; P_{k|k-1}^{(2)} \otimes \mathrm{cst}(R_k^{(2)})^T\, B_{k_4}^T \\ &\quad + H_k\, E[\mathbf{v}_k^T \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1}^T \otimes \mathbf{v}_k^T]\, B_{k_6}^T \;+\; \mathrm{cst}(R_k^{(2)})^T \otimes P_{k|k-1}^{(2)}\, B_{k_7}^T \\ &\quad + \mathrm{cst}(P_{k|k-1}^{(2)})^T \otimes R_k^{(2)}\, B_{k_2}^T + E[\tilde{\mathbf{x}}_{k|k-1}^T \otimes \mathbf{v}_k^T \otimes \mathbf{v}_k^T \otimes \tilde{\mathbf{x}}_{k|k-1}^T]\, B_{k_3}^T \\ &\quad + R_k^{(2)} \otimes \mathrm{cst}(P_{k|k-1}^{(2)})^T\, B_{k_5}^T + \mathrm{scst}(R_k^{(4)})^T \end{aligned} \tag{3.51}$$

$$
\begin{aligned}
E[\tilde{\alpha}_k \tilde{\alpha}_k^T] = {} & B_{k_1} P^{(6)}_{k|k-1} B^T_{k_1} + B_{k_4} \mathrm{scst}(P^{(4)}_{k|k-1})^T \otimes \mathrm{cst}(R^{(2)}_k) B^T_{k_1} \\
& + B_{k_6} C_{k_1} B^T_{k_1} + B_{k_7} \mathrm{cst}(R^{(2)}_k) \otimes \mathrm{scst}(P^{(4)}_{k|k-1})^T B^T_{k_1} \\
& + B_{k_2} P^{(4)}_{k|k-1} \otimes R^{(2)}_k B^T_{k_2} + B_{k_3} C_{k_2} B^T_{k_2} \\
& + B_{k_5} C_{k_3} B^T_{k_2} + B_{k_8} \mathrm{cst}(P^{(2)}_{k|k-1})^T \otimes \mathrm{scst}(R^{(4)}_k)^T B^T_{k_2} \\
& + B_{k_2} C_{k_4} B^T_{k_3} + B_{k_3} C_{k_5} B^T_{k_3} + B_{k_5} C_{k_6} B^T_{k_3} + B_{k_8} C_{k_7} B^T_{k_3} \\
& + B_{k_1} \mathrm{scst}(P^{(4)}_{k|k-1}) \otimes \mathrm{cst}(R^{(2)}_k)^T B^T_{k_4} + B_{k_4} P^{(2)}_{k|k-1} \otimes R^{(4)}_k B^T_{k_4} \\
& + B_{k_6} C_{k_8} B^T_{k_4} + B_{k_7} C_{k_9} B^T_{k_4} + B_{k_2} C_{k_{10}} B^T_{k_5} + B_{k_3} C_{k_{11}} B^T_{k_5} \qquad (3.52) \\
& + B_{k_5} R^{(2)}_k \otimes P^{(4)}_{k|k-1} B^T_{k_5} + B_{k_8} \mathrm{scst}(R^{(4)}_k) \otimes \mathrm{cst}(P^{(2)}_{k|k-1})^T B^T_{k_5} \\
& + B_{k_1} C_{k_{12}} B^T_{k_6} + B_{k_4} C_{k_{13}} B^T_{k_6} + B_{k_6} C_{k_{14}} B^T_{k_6} + B_{k_7} C_{k_{15}} B^T_{k_6} \\
& + B_{k_1} \mathrm{cst}(R^{(2)}_k)^T \otimes \mathrm{scst}(P^{(4)}_{k|k-1}) B^T_{k_7} + B_{k_4} C_{k_{16}} B^T_{k_7} \\
& + B_{k_6} C_{k_{17}} B^T_{k_7} + B_{k_7} R^{(4)}_k \otimes P^{(2)}_{k|k-1} B^T_{k_7} \\
& + B_{k_2} \mathrm{cst}(P^{(2)}_{k|k-1}) \otimes \mathrm{scst}(R^{(4)}_k)^T B^T_{k_8} + B_{k_3} C_{k_{18}} B^T_{k_8} \\
& + B_{k_5} \mathrm{scst}(R^{(4)}_k)^T \otimes \mathrm{cst}(P^{(2)}_{k|k-1}) B^T_{k_8} + B_{k_8} R^{(6)}_k B^T_{k_8}
\end{aligned}
$$

The parameters C_{k_i} are 6^{th} order functions of expectations of the measure-

ment and prediction errors. They are defined by

$$
\begin{aligned}
C_{k_1} &= E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}_k] \\
C_{k_2} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k] \\
C_{k_3} &= E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k] \\
C_{k_4} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_5} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_6} &= E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_7} &= E[\mathbf{v}_k \otimes \mathbf{v}_k \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_8} &= E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k \otimes \mathbf{v}^T_k] \\
C_{k_9} &= E[\mathbf{v}_k \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k \otimes \mathbf{v}^T_k] \\
C_{k_{10}} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_{11}} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_{12}} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k] \\
C_{k_{13}} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}_k \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k] \\
C_{k_{14}} &= E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k] \\
C_{k_{15}} &= E[\mathbf{v}_k \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1} \otimes \mathbf{v}^T_k] \\
C_{k_{16}} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}_k \otimes \mathbf{v}^T_k \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_{17}} &= E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}^T_k \otimes \mathbf{v}^T_k \otimes \tilde{\mathbf{x}}^T_{k|k-1}] \\
C_{k_{18}} &= E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}^T_k \otimes \mathbf{v}^T_k \otimes \mathbf{v}^T_k]
\end{aligned}
\tag{3.53}
$$

The remaining work involves evaluation of the prediction error moments and

the filter error moments. The prediction error, $\tilde{\mathbf{x}}_{k|k-1}$ is given by (3.29). The 2^{nd} and 4^{th} prediction moments are generated from the prediction errors. These moments are expressed as

$$\begin{aligned} P^{(2)}_{k|k-1} &= E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}^T_{k|k-1}] \\ &= \Phi_{k-1} P^{(2)}_{k-1|k-1} \Phi^T_{k-1} + Q^{(2)}_{k-1} \end{aligned} \tag{3.54}$$

$$\begin{aligned} P^{(4)}_{k|k-1} &= E[(\tilde{\mathbf{x}}^{\otimes 2}_{k|k-1})\,(\tilde{\mathbf{x}}^{\otimes 2}_{k|k-1})^T] \\ &= \Phi_{k-1} \otimes \Phi_{k-1}\; P^{(4)}_{k-1|k-1}\; \Phi^T_{k-1} \otimes \Phi^T_{k-1} \;+\; Q^{(4)}_{k-1} \\ &+\; \Phi_{k-1} \otimes I_n\; P^{(2)}_{k-1|k-1} \otimes Q^{(2)}_{k-1}\; \Phi^T_{k-1} \otimes I_n \\ &+\; \Phi_{k-1} \otimes \Phi_{k-1}\; \mathrm{cst}(P^{(2)}_{k-1|k-1})\; \mathrm{cst}(Q^{(2)}_{k-1})^T \\ &+\; \Phi_{k-1} \otimes I_n\; E[\tilde{\mathbf{x}}_{k-1|k-1} \otimes \mathbf{w}^T_{k-1} \otimes \mathbf{w}_{k-1} \otimes \tilde{\mathbf{x}}_{k-1|k-1}]\; I_n \otimes \Phi^T_{k-1} \\ &+\; I_n \otimes \Phi_{k-1}\; E[\mathbf{w}_{k-1} \otimes \tilde{\mathbf{x}}^T_{k-1|k-1} \otimes \tilde{\mathbf{x}}_{k-1|k-1} \otimes \mathbf{w}_{k-1}]\; \Phi^T_{k-1} \otimes I_n \\ &+\; \mathrm{cst}(Q^{(2)}_{k-1})\; \mathrm{cst}(P^{(2)}_{k-1|k-1})^T\; \Phi^T_{k-1} \otimes \Phi^T_{k-1} \\ &+\; I_n \otimes \Phi_{k-1}\; Q^{(2)}_{k-1} \otimes P^{(2)}_{k-1|k-1}\; I_n \otimes \Phi^T_{k-1} \end{aligned} \tag{3.55}$$

Similarly, using equation (3.37) the moments of the filter error can be evaluated. Let

$$A_k \equiv (I_n - K^{(1)}_k H_k) \tag{3.56}$$

Then the filter variance becomes

$$
\begin{aligned}
P_{k|k}^{(2)} &= E[\tilde{\mathbf{x}}_{k|k}\tilde{\mathbf{x}}_{k|k}^T] \\
&= A_k\, P_{k|k-1}^{(2)}\, A_k^T \;+\; K_k^{(1)} R_k^{(2)} K_k^{(1)\,T} \\
&- A_k \mathrm{scst}(P_{k|k-1}^{(4)})^T B_{k_1}^T K_k^{(3)\,T} - K_k^{(3)} B_{k_1} \mathrm{scst}(P_{k|k-1}^{(4)}) A_k^T \\
&- A_k P_{k|k-1}^{(2)} \otimes \mathrm{cst}(R_k^{(2)})^T B_{k_4}^T K_k^{(3)\,T} - K_k^{(3)} B_{k_4} \mathrm{cst}(R_k^{(2)}) \otimes P_{k|k-1}^{(2)} A_k^T \\
&- A_k E[\tilde{\mathbf{x}}_{k|k-1}^T \otimes \mathbf{v}_k \otimes \mathbf{v}_k^T \otimes \tilde{\mathbf{x}}_{k|k-1}^T] B_{k_6}^T K_k^{(3)\,T} \\
&- K_k^{(3)} B_{k_6} E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k \otimes \mathbf{v}_k^T \otimes \tilde{\mathbf{x}}_{k|k-1}] A_k^T \\
&- A_k \mathrm{cst}(R_k^{(2)})^T \otimes P_{k|k-1}^{(2)} B_{k_7}^T K_k^{(3)\,T} - K_k^{(3)} B_{k_7} P_{k|k-1}^{(2)} \otimes \mathrm{cst}(R_k^{(2)}) A_k^T \\
&+ K_k^{(1)} \mathrm{cst}(P_{k|k-1}^{(2)})^T \otimes R_k^{(2)} B_{k_2}^T K_k^{(3)\,T} + K_k^{(3)} B_{k_2} R_k^{(2)} \otimes \mathrm{cst}(P_{k|k-1}^{(2)}) K_k^{(1)\,T} \\
&+ K_k^{(1)} E[\mathbf{v}_k^T \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1}^T \otimes \mathbf{v}_k^T] B_{k_3}^T K_k^{(3)\,T} \\
&+ K_k^{(3)} B_{k_3} E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \tilde{\mathbf{x}}_{k|k-1}^T \otimes \mathbf{v}_k] K_k^{(1)\,T} \\
&+ K_k^{(1)} R_k^{(2)} \otimes \mathrm{cst}(P_{k|k-1}^{(2)})^T B_{k_5}^T K_k^{(3)\,T} + K_k^{(3)} B_{k_5} \mathrm{cst}(P_{k|k-1}^{(2)}) \otimes R_k^{(2)} K_k^{(1)\,T} \\
&+ K_k^{(1)} \mathrm{scst}(R_k^{(4)})^T K_k^{(3)\,T} + K_k^{(3)} \mathrm{scst}(R_k^{(4)}) K_k^{(1)\,T}
\end{aligned}
\tag{3.57}
$$

Equation (3.37) dictates that the filter variance should include 6^{th} order functions of the prediction and measurement error moments. These higher order terms are not included in the filter variance expression, just as all 5^{th} and higher order terms are disregarded in the development of the asymmetrical filter. By doing so it is implicitly assumed that the contributions from these higher order terms are negligible. As noted previously, if these terms were included in the derivation of the filter moments it would necessitate some approximation procedure for these high order prediction and measurement moments, since only 2^{nd} and 4^{th} order prediction moments are propagated. Similarly the 4^{th} order filter moment requires the availability of 6^{th}, 8^{th}, 10^{th}, and 12^{th} order prediction and measurement error moments. Again the 4^{th} order expansion is truncated to include only 4^{th} order functions of the

prediction and measurement errors. The resulting moment equation becomes

$$
\begin{aligned}
P_{k|k}^{(4)} &= E[(\tilde{\mathbf{x}}_{k|k}^{\otimes 2})\,(\tilde{\mathbf{x}}_{k|k}^{\otimes 2})^T] \\
&= A_k \otimes A_k\; P_{k|k-1}^{(4)}\; A_k^T \otimes A_k^T \\
&+\; A_k \otimes K_k^{(1)}\; P_{k|k-1}^{(2)} \otimes R_k^{(2)}\; A_k^T \otimes K_k^{(1)\,T} \\
&+\; A_k \otimes A_k\; \mathrm{cst}(P_{k|k-1}^{(2)})\; \mathrm{cst}(R_k^{(2)})^T\; K_k^{(1)\,T} \otimes K_k^{(1)\,T} \\
&+\; A_k \otimes K_k^{(1)}\; E[\tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k^T \otimes \mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1}^T]\; K_k^{(1)\,T} \otimes A_k^T \\
&+\; K_k^{(1)} \otimes A_k\; E[\mathbf{v}_k \otimes \tilde{\mathbf{x}}_{k|k-1}^T \otimes \tilde{\mathbf{x}}_{k|k-1} \otimes \mathbf{v}_k^T]\; A_k^T \otimes K_k^{(1)\,T} \\
&+\; K_k^{(1)} \otimes K_k^{(1)}\; \mathrm{cst}(R_k^{(2)})\; \mathrm{cst}(P_{k|k-1}^{(2)})^T\; A_k^T \otimes A_k^T \\
&+\; K_k^{(1)} \otimes A_k\; R_k^{(2)} \otimes P_{k|k-1}^{(2)}\; K_k^{(1)\,T} \otimes A_k^T \\
&+\; K_k^{(1)} \otimes K_k^{(1)}\; R_k^{(4)}\; K_k^{(1)\,T} \otimes K_k^{(1)\,T}
\end{aligned}
\tag{3.58}
$$

The filter equations (3.38, 3.57, 3.58), gain equations (3.44), and prediction equations (3.28, 3.54, 3.55) constitute the discrete filter relations for non-Gaussian noise with arbitrary symmetrical distributions. Similar to the asymmetrical filter, these relations are suboptimal in that they do not completely characterize the noise distributions since they make use of only the first four moments of the distributions.

3.4 Nonlinear Non-Gaussian Filtering

It is straightforward to extend the high order filters that have been derived for linear plant and measurement models to nonlinear plant and nonlinear measurement equations in non-Gaussian noise. Using linearized models based on 1^{st} order Taylor series expansions, replace Φ_{k-1} and H_k in the linear model non-Gaussian filters with

$$
F_k = \left.\frac{\partial \mathbf{f}_k(\mathbf{x}_k}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k-1|k-1}}
$$

$$
H_k = \left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}
$$

where $\mathbf{f}_k(\mathbf{x}_k)$ and $\mathbf{h}_k(\mathbf{x}_k)$ are the system and measurement nonlinearities. The equations for the nonlinear non-Gaussian filter have the same form as those for the linear non-Gaussian filter. However the nonlinear filter requires the computation of F_k and H_k at every iteration.

Likewise, the locally iterated Kalman filter, which is discussed in Section 2.4.3, can be obtained from the non-Gaussian filter equations. The locally iterated non-Gaussian filter requires the computation of all the filtered estimates, all the gains, and all the moments of the a *posteriori* density function on every iteration in each step in the filtering process.

3.5 Non-Gaussian Filters for Scalar Models

In this section the scalar equations for the symmetrical and asymmetrical filters are presented. It is shown that these filters reduce to the Kalman filter equations for the special case of Gaussian noise. In contrast to the development of the vector-based filter equations in which the equations are truncated in order to reduce the filter complexity, the filter equations for the scalar models are derived without truncation.

The scalar state and measurement equations are given by

$$\begin{aligned} x_k &= \phi_{k-1} x_{k-1} + w_{k-1} \\ z_k &= h_k x_k + v_k \end{aligned} \tag{3.59}$$

The predicted estimate is $\hat{x}_{k|k-1} = \phi_{k-1}\hat{x}_{k-1|k-1}$ and the prediction error $\tilde{x}_{k|k-1} = x_k - \hat{x}_{k|k-1}$ is given by

$$\tilde{x}_{k|k-1} = \phi_{k-1}\tilde{x}_{k-1|k-1} + w_{k-1} \tag{3.60}$$

3.5.1 Scalar Asymmetrical Filter

The scalar equations for the asymmetrical filter are presented in this section. These equations have the same form as the vector equations given previously with the exception that the moment equations are not truncated. When the measurement noise, process noise, or initial estimation error has asymmetrical distributions the filter equation for the scalar model is obtained from (3.14)

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + k_k^{(1)}\tilde{z}_k + k_k^{(2)}\alpha_k \tag{3.61}$$

where

$$\alpha_k = \tilde{z}_k^2 - h_k^2 p_{k|k-1}^{(2)} - r_k^{(2)} \tag{3.62}$$

The filter error $\tilde{x}_{k|k} = x_k - \hat{x}_{k|k}$ becomes

$$\tilde{x}_{k|k} = \tilde{x}_{k|k-1} - k_k^{(1)}\tilde{z}_k - k_k^{(2)}\alpha_k \tag{3.63}$$

The scalar filter gains (3.20) become

$$\begin{aligned} k_k^{(1)} &= \frac{E[\tilde{x}_{k|k-1}\tilde{z}_k]E[\alpha_k^2] - E[\tilde{x}_{k|k-1}\alpha_k]E[\tilde{z}_k\alpha_k]}{E[\tilde{z}_k^2]E[\alpha_k^2] - E[\tilde{z}_k\alpha_k]^2} \\ k_k^{(2)} &= \frac{E[\tilde{x}_{k|k-1}\alpha_k]E[\tilde{z}_k^2] - E[\tilde{x}_{k|k-1}\tilde{z}_k]E[\tilde{z}_k\alpha_k]}{E[\alpha_k^2]E[\tilde{z}_k^2] - E[\tilde{z}_k\alpha_k]^2} \end{aligned} \tag{3.64}$$

where

$$E[\tilde{x}_{k|k-1}\tilde{z}_k] = h_k p^{(2)}_{k|k-1}$$

$$E[\alpha_k \tilde{x}_{k|k-1}] = h_k^2 p^{(3)}_{k|k-1}$$

$$E[\alpha_k \tilde{z}_k] = h_k^3 p^{(3)}_{k|k-1} + r_k^{(3)} \tag{3.65}$$

$$E[\tilde{z}_k^2] = h_k^2 p^{(2)}_{k|k-1} + r_k^{(2)}$$

$$E[\alpha_k^2] = -h_k^4 {p^{(2)}_{k|k-1}}^2 + h_k^4 p^{(4)}_{k|k-1} + 4h_k^2 r_k^{(2)} p^{(2)}_{k|k-1} - {r_k^{(2)}}^2 + r_k^{(4)}$$

The scalar versions of the prediction moments obtained from (3.30- 3.32) are given by

$$p^{(2)}_{k|k-1} = \phi^2_{k-1} p^{(2)}_{k-1|k-1} + q^{(2)}_{k-1} \tag{3.66}$$

$$p^{(3)}_{k|k-1} = \phi^3_{k-1} p^{(3)}_{k-1|k-1} + q^{(3)}_{k-1} \tag{3.67}$$

$$p^{(4)}_{k|k-1} = \phi^4_{k-1} p^{(4)}_{k-1|k-1} + 6\phi^2_{k-1} q^{(2)}_{k-1} p^{(2)}_{k-1|k-1} + q^{(4)}_{k-1} \tag{3.68}$$

The 2nd moment $p^{(2)}_{k|k} = E[\tilde{x}^2_{k|k}]$ can be partitioned into $2^{nd}, 3^{rd}$, and 4^{th} order components of the measurement and prediction moments. Let

$$p^{(2)}_{k|k} = p^{(2)}_{k|k_2} + p^{(2)}_{k|k_3} + p^{(2)}_{k|k_4} \tag{3.69}$$

where $p^{(2)}_{k|k_i}$ consists of i^{th} order moments of the measurement and prediction error. The $p^{(2)}_{k|k_i}$ are given by

$$p^{(2)}_{k|k_2} = a_k^2 p^{(2)}_{k|k-1} + [k_k^{(1)}]^2 r_k^{(2)} \tag{3.70}$$

$$p^{(2)}_{k|k_3} = 2k_k^{(1)} k_k^{(2)} r_k^{(3)} - 2a_k k_k^{(2)} h_k^2 p^{(3)}_{k|k-1} \tag{3.71}$$

$$p^{(2)}_{k|k_4} = [k^{(2)}_k]^2(h^4_k(p^{(4)}_{k|k-1} - [p^{(2)}_{k|k-1}]^2) + 4h^2_k p^{(2)}_{k|k-1} r^{(2)}_k - [r^{(2)}_k]^2 + r^{(4)}_k) \tag{3.72}$$

where $a_k = 1 - h_k k^{(1)}_k$.

The 3rd moment $p^{(3)}_{k|k} = E[\tilde{x}^3_{k|k}]$ can be partitioned into $3^{th}, 4^{th}, 5^{th}$, and 6^{th} order components of the measurement and prediction moments. Let

$$p^{(3)}_{k|k} = p^{(3)}_{k|k_3} + p^{(3)}_{k|k_4} + p^{(3)}_{k|k_5} + p^{(3)}_{k|k_6} \tag{3.73}$$

where $p^{(3)}_{k|k_i}$ consists of i^{th} order components of the measurement and prediction moments. The $p^{(3)}_{k|k_i}$ are given by

$$p^{(3)}_{k|k_3} = a^3_k p^{(3)}_{k|k-1} - [k^{(1)}_k]^3 r^{(3)}_k \tag{3.74}$$

$$\begin{aligned} p^{(3)}_{k|k_4} = {} & a^2_k k^{(2)}_k h^2_k (3[p^{(2)}_{k|k-1}]^2 - 3p^{(4)}_{k|k-1}) + [k^{(1)}_k]^2 k^{(2)}_k (3[r^{(2)}_k]^2 - 3r^{(4)}_k) \\ & + 12a_k k^{(1)}_k k^{(2)}_k h_k p^{(2)}_{k|k-1} r^{(2)}_k \end{aligned} \tag{3.75}$$

$$\begin{aligned} p^{(3)}_{k|k_5} = {} & a_k [k^{(2)}_k]^2 (h^4_k (3p^{(5)}_{k|k-1} - 6p^{(2)}_{k|k-1} p^{(3)}_{k|k-1}) + 12h^2_k p^{(3)}_{k|k-1} r^{(2)}_k \\ & + 12h_k p^{(2)}_{k|k-1} r^{(3)}_k) + k^{(1)}_k [k^{(2)}_k]^2 (-12h^3_k p^{(3)}_{k|k-1} r^{(2)}_k \\ & - 12h^2_k p^{(2)}_{k|k-1} r^{(3)}_k + 6r^{(2)}_k r^{(3)}_k - 3r^{(5)}_k) \end{aligned} \tag{3.76}$$

$$\begin{aligned} p^{(3)}_{k|k_6} = {} & [k^{(2)}_k]^3 (h^6_k (-2[p^{(2)}_{k|k-1}]^3 + 3p^{(2)}_{k|k-1} p^{(4)}_{k|k-1} - p^{(6)}_{k|k-1}) \\ & + h^4_k (12[p^{(2)}_{k|k-1}]^2 r^{(2)}_k - 12p^{(4)}_{k|k-1} r^{(2)}_k) - 20h^3_k p^{(3)}_{k|k-1} r^{(3)}_k \\ & + h^2_k p^{(2)}_{k|k-1} (12[r^{(2)}_k]^2 - 12r^{(4)}_k) - 2[r^{(2)}_k]^3 + 3r^{(2)}_k r^{(4)}_k - r^{(6)}_k) \end{aligned} \tag{3.77}$$

The 4th moment $p^{(4)}_{k|k} = E[\tilde{x}^4_{k|k}]$ can be partitioned into $4^{th}, 5^{th}, \cdots, 8^{th}$ order components of the measurements and prediction moments. Let

$$p^{(4)}_{k|k} = p^{(4)}_{k|k_4} + p^{(4)}_{k|k_5} + p^{(4)}_{k|k_6} + p^{(4)}_{k|k_7} + p^{(4)}_{k|k_8} \tag{3.78}$$

where $p^{(4)}_{k|k_i}$ consists of i^{th} order components of the measurement and prediction moments. The $p^{(4)}_{k|k_i}$ are given by

$$p^{(4)}_{k|k_4} = a_k^4 p^{(4)}_{k|k-1} + 6a_k^2[k_k^{(1)}]^2 p^{(2)}_{k|k-1} r_k^{(2)} + [k_k^{(1)}]^4 r_k^{(4)} \tag{3.79}$$

$$\begin{aligned} p^{(4)}_{k|k_5} = {} & 4a_k^3 k_k^{(2)} h_k^2 (p^{(2)}_{k|k-1} p^{(3)}_{k|k-1} - p^{(5)}_{k|k-1}) \\ & + 12a_k^2 k_k^{(1)} k_k^{(2)} (2h_k p^{(3)}_{k|k-1} r_k^{(2)} + p^{(2)}_{k|k-1} r_k^{(3)}) \\ & + 12a_k [k_k^{(1)}]^2 k_k^{(2)} (-h_k^2 p^{(3)}_{k|k-1} r_k^{(2)} - 2h_k p^{(2)}_{k|k-1} r_k^{(3)}) \\ & + 4[k_k^{(1)}]^3 k_k^{(2)} (r_k^{(5)} - r_k^{(2)} r_k^{(3)}) \end{aligned} \tag{3.80}$$

$$\begin{aligned} p^{(4)}_{k|k_6} = {} & a_k^2 [k_k^{(2)}]^2 (6h_k^4 ([p^{(2)}_{k|k-1}]^3 - 12 p^{(2)}_{k|k-1} p^{(4)}_{k|k-1} + 6p^{(6)}_{k|k-1}) \\ & + 24h_k^2 p^{(4)}_{k|k-1} r_k^{(2)} + 24h_k p^{(3)}_{k|k-1} r_k^{(3)} + 6p^{(2)}_{k|k-1} (r_k^{(4)} - [r_k^{(2)}]^2)) \\ & + a_k k_k^{(1)} [k_k^{(2)}]^2 (48h_k^3 r_k^{(2)} ([p^{(2)}_{k|k-1}]^2 - p^{(4)}_{k|k-1}) - 72h_k^2 p^{(3)}_{k|k-1} r_k^{(3)} \\ & + 48h_k p^{(2)}_{k|k-1} ([r_k^{(2)}]^2 - r_k^{(4)})) \\ & + [k_k^{(1)}]^2 [k_k^{(2)}]^2 (6h_k^4 (p^{(4)}_{k|k-1} r_k^{(2)} - [p^{(2)}_{k|k-1}]^2 r_k^{(2)}) + 24h_k^3 p^{(3)}_{k|k-1} r_k^{(3)} \\ & + 24h_k^2 p^{(2)}_{k|k-1} r_k^{(4)} + 6[r_k^{(2)}]^3 - 12 r_k^{(2)} r_k^{(4)} + 6r_k^{(6)}) \end{aligned} \tag{3.81}$$

$$
\begin{aligned}
p^{(4)}_{k|k_7} &= a_k[k^{(2)}_k]^3(h^6_k(-12[p^{(2)}_{k|k-1}]^2p^{(3)}_{k|k-1} + 12p^{(2)}_{k|k-1}p^{(5)}_{k|k-1} - 4p^{(7)}_{k|k-1}) \\
&+ h^4_k(48p^{(2)}_{k|k-1}p^{(3)}_{k|k-1}r^{(2)}_k - 48p^{(5)}_{k|k-1}r^{(2)}_k) + h^3_k(48[p^{(2)}_{k|k-1}]^2r^{(3)}_k - 80p^{(4)}_{k|k-1}r^{(3)}_k) \\
&+ h^2_kp^{(3)}_{k|k-1}(60[r^{(2)}_k]^2 - 60r^{(4)}_k) + h_kp^{(2)}_{k|k-1}(48r^{(2)}_kr^{(3)}_k - 24r^{(5)}_k)) \\
&+ k^{(1)}_k[k^{(2)}_k]^3(h^5_k(24p^{(5)}_{k|k-1}r^{(2)}_k - 48p^{(2)}_{k|k-1}p^{(3)}_{k|k-1}r^{(2)}_k) \\
&\quad + h^4_k(60p^{(4)}_{k|k-1}r^{(3)}_k - 60[p^{(2)}_{k|k-1}]^2r^{(3)}_k) + h^3_kp^{(3)}_{k|k-1}(80r^{(4)}_k - 48[r^{(2)}_k]^2) \\
&\quad + h^2_kp^{(2)}_{k|k-1}(48r^{(5)}_k - 48r^{(2)}_kr^{(3)}_k) + 12[r^{(2)}_k]^2r^{(3)}_k - 12r^{(2)}_kr^{(5)}_k + 4r^{(7)}_k)
\end{aligned}
\tag{3.82}
$$

$$
\begin{aligned}
p^{(4)}_{k|k_8} &= [k^{(2)}_k]^4(h^8_k(-3[p^{(2)}_{k|k-1}]^4 + 6[p^{(2)}_{k|k-1}]^2p^{(4)}_{k|k-1} - 4p^{(2)}_{k|k-1}p^{(6)}_{k|k-1} + p^{(8)}_{k|k-1}) \\
&+ h^6_k(24[p^{(2)}_{k|k-1}]^3r^{(2)}_k - 48p^{(2)}_{k|k-1}p^{(4)}_{k|k-1}r^{(2)}_k + 24p^{(6)}_{k|k-1}r^{(2)}_k) \\
&+ h^5_k(56p^{(5)}_{k|k-1}r^{(3)}_k - 80p^{(2)}_{k|k-1}p^{(3)}_{k|k-1}r^{(3)}_k) \\
&+ h^4_k([p^{(2)}_{k|k-1}]^2(54[r^{(2)}_k]^2 - 54r^{(4)}_k) + p^{(4)}_{k|k-1}(70r^{(4)}_k - 54[r^{(2)}_k]^2)) \\
&+ h^3_kp^{(3)}_{k|k-1}(56r^{(5)}_k - 80r^{(2)}_kr^{(3)}_k) + h^2_kp^{(2)}_{k|k-1}(24[r^{(2)}_k]^3 - 48r^{(2)}_kr^{(4)}_k + 24r^{(6)}_k) \\
&- 3[r^{(2)}_k]^4 + 6[r^{(2)}_k]^2r^{(4)}_k - 4r^{(2)}_kr^{(6)}_k + r^{(8)}_k)
\end{aligned}
\tag{3.83}
$$

3.5.1.1 Special Case - Symmetrical Distributions

In the case where the initial estimation error, the measurement noise, and the

process noise all have symmetrical distributions the expectations in (3.59) reduce to

$$\begin{aligned} E[\tilde{x}_{k|k-1}\tilde{z}_k] &= h_k p^{(2)}_{k|k-1} \\ E[\alpha_k \tilde{x}_{k|k-1}] &= 0 \\ E[\alpha_k \tilde{z}_k] &= 0 \\ E[\tilde{z}_k^2] &= h_k^2 p^{(2)}_{k|k-1} + r_k^{(2)} \\ E[\alpha_k^2] &= -h_k^4 [p^{(2)}_{k|k-1}]^2 + h_k^4 p^{(4)}_{k|k-1} + 4h_k^2 r_k^{(2)} p^{(2)}_{k|k-1} - [r_k^{(2)}]^2 + r_k^{(4)} \end{aligned} \tag{3.84}$$

Substituting into (3.58) the filter gains become

$$\begin{aligned} k_k^{(1)} &= E[\tilde{x}_{k|k-1}\tilde{z}_k]\; E[\tilde{z}_k^2] \\ k_k^{(2)} &= 0 \end{aligned} \tag{3.85}$$

It is observed that $k_k^{(1)}$ is now the standard Kalman filter gain. With $k_k^{(2)} = 0$ the 3^{rd} and 4^{th} order components of the 2^{nd} filter moment (3.69) are zero. Thus there is no need to compute 3^{rd} and higher order filter moments. The resulting equations are the standard Kalman filter equations for linear systems.

3.5.2 Scalar Symmetrical Filter

The scalar filter equations for the symmetrical non-Gaussian are obtained in the same manner as the vector equations with the exception that the moment equations are not truncated. From (3.38) the filter equation becomes

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + k_k^{(1)}\tilde{z}_k + k_k^{(3)}(\tilde{z}_k^3) \tag{3.86}$$

The filter error $\tilde{x}_{k|k} = x_k - \hat{x}_{k|k}$ becomes

$$\tilde{x}_{k|k} = \tilde{x}_{k|k-1} - k_k^{(1)}\tilde{z}_k - k_k^{(3)}(\tilde{z}_k^3) \tag{3.87}$$

The scalar filter gains are obtained from (3.44). They are given by

$$k_k^{(1)} = \frac{E[\tilde{x}_{k|k-1}\tilde{z}_k]E[\tilde{z}_k^6] - E[\tilde{x}_{k|k-1}\tilde{z}_k^3]E[\tilde{z}_k^4]}{E[\tilde{z}_k^2]E[\tilde{z}_k^6] - E[\tilde{z}_k^4]^2}$$
$$k_k^{(3)} = \frac{E[\tilde{x}_{k|k-1}\tilde{z}_k^3]E[\tilde{z}_k^2] - E[\tilde{x}_{k|k-1}\tilde{z}_k]E[\tilde{z}_k^4]}{E[\tilde{z}_k^6]E[\tilde{z}_k^2] - E[\tilde{z}_k^4]^2} \tag{3.88}$$

where

$$E[\tilde{x}_{k|k-1}\tilde{z}_k] = h_k p_{k|k-1}^{(2)}$$
$$E[\tilde{z}_k^2] = h_k^2 p_{k|k-1}^{(2)} + r_k^{(2)}$$
$$E[\tilde{x}_{k|k-1}\tilde{z}_k^3] = h_k^3 p_{k|k-1}^{(4)} + 3h_k r_k^{(2)} p_{k|k-1}^{(2)} \tag{3.89}$$
$$E[\tilde{z}_k^4] = h_k^4 p_{k|k-1}^{(4)} + 6h_k^2 r_k^{(2)} p_{k|k-1}^{(2)} + r_k^{(4)}$$
$$E[\tilde{z}_k^6] = h_k^6 p_{k|k-1}^{(6)} + 15h_k^4 r_k^{(2)} p_{k|k-1}^{(4)} + 15h_k^2 r_k^{(4)} p_{k|k-1}^{(2)} + r_k^{(6)}$$

The prediction moments are scalar versions of (3.54) and (3.55)

$$p_{k|k-1}^{(2)} = \phi_{k-1}^2 p_{k-1|k-1}^{(2)} + q_{k-1}^{(2)} \tag{3.90}$$

$$p_{k|k-1}^{(4)} = \phi_{k-1}^4 p_{k-1|k-1}^{(4)} + 6\phi_{k-1}^2 q_{k-1}^{(2)} p_{k-1|k-1}^{(2)} + q_{k-1}^{(4)} \tag{3.91}$$

The filter variance $p_{k|k}^{(2)} = E[\tilde{x}_{k|k}^2]$ can be partitioned into $2^{nd}, 4^{th}$, and 6^{th} order components of the measurement and prediction moments. Let

$$p_{k|k}^{(2)} = p_{k|k_2}^{(2)} + p_{k|k_4}^{(2)} + p_{k|k_6}^{(2)} \tag{3.92}$$

where $p_{k|k_i}^{(2)}$ consists of i^{th} order components of the measurement and prediction moments. The $p_{k|k_i}^{(2)}$ are given by

$$p^{(2)}_{k|k_2} = a_k^2 p^{(2)}_{k|k-1} + [k_k^{(1)}]^2 r_k^{(2)} \tag{3.93}$$

$$p^{(2)}_{k|k_4} = a_k k_k^{(3)}(-2h_k^3 p^{(4)}_{k|k-1} - 6h_k p^{(2)}_{k|k-1} r_k^{(2)}) + k_k^{(1)} k_k^{(3)}(6h_k^2 p^{(2)}_{k|k-1} r_k^{(2)} + 2r_k^{(4)}) \tag{3.94}$$

$$p^{(2)}_{k|k_6} = [k_k^{(3)}]^2(h_k^6 p^{(6)}_{k|k-1} + 15h_k^4 p^{(4)}_{k|k-1} r_k^{(2)} + 15h_k^2 p^{(2)}_{k|k-1} r_k^{(4)} + r_k^{(6)}) \tag{3.95}$$

where $a_k = 1 - h_k k_k^{(1)}$.

The 4th moment $p^{(4)}_{k|k} = E[\tilde{x}^4_{k|k}]$ can be partitioned into $4^{th}, 6^{th}, \cdots, 12^{th}$ order components of the measurements and prediction moments. Let

$$p^{(4)}_{k|k} = p^{(4)}_{k|k_4} + p^{(4)}_{k|k_6} + p^{(4)}_{k|k_8} + p^{(4)}_{k|k_{10}} + p^{(4)}_{k|k_{12}} \tag{3.96}$$

where $p^{(4)}_{k|k_i}$ consists of i^{th} order components of the measurement and prediction moments. The $p^{(4)}_{k|k_i}$ are given by

$$p^{(4)}_{k|k_4} = a_k^4 p^{(4)}_{k|k-1} + 6a_k^2 [k_k^{(1)}]^2 p^{(2)}_{k|k-1} r_k^{(2)} + [k_k^{(1)}]^4 r_k^{(4)} \tag{3.97}$$

$$\begin{aligned} p^{(4)}_{k|k_6} &= a_k^3 k_k^{(3)}(-4h_k^3 p^{(6)}_{k|k-1} - 12h_k p^{(4)}_{k|k-1} r_k^{(2)}) \\ &\quad + a_k^2 k_k^{(1)} k_k^{(3)}(36h_k^2 p^{(4)}_{k|k-1} r_k^{(2)} + 12p^{(2)}_{k|k-1} r_k^{(4)}) \\ &\quad + a_k [k_k^{(1)}]^2 k_k^{(3)}(-12h_k^3 p^{(4)}_{k|k-1} r_k^{(2)} - 36h_k p^{(2)}_{k|k-1} r_k^{(4)}) \\ &\quad + [k_k^{(1)}]^3 k_k^{(3)}(12h_k^2 p^{(2)}_{k|k-1} r_k^{(4)} + 4r_k^{(6)}) \end{aligned} \tag{3.98}$$

$$\begin{aligned} p^{(4)}_{k|k_8} &= a_k^2 [k_k^{(3)}]^2(6h_k^6 p^{(8)}_{k|k-1} + 90h_k^4 p^{(6)}_{k|k-1} r_k^{(2)} + 90h_k^2 p^{(4)}_{k|k-1} r_k^{(4)} + 6p^{(2)}_{k|k-1} r_k^{(6)}) \\ &\quad + a_k k_k^{(1)} [k_k^{(3)}]^2(-72h_k^5 p^{(6)}_{k|k-1} r_k^{(2)} - 240h_k^3 p^{(4)}_{k|k-1} r_k^{(4)} - 72h_k p^{(2)}_{k|k-1} r_k^{(6)}) \\ &\quad + [k_k^{(1)}]^2 [k_k^{(3)}]^2(6h_k^6 p^{(6)}_{k|k-1} r_k^{(2)} + 90h_k^4 p^{(4)}_{k|k-1} r_k^{(4)} + 90h_k^2 p^{(2)}_{k|k-1} r_k^{(6)} + 6r_k^{(8)}) \end{aligned} \tag{3.99}$$

$$
\begin{aligned}
p^{(4)}_{k|k_{10}} &= a_k[k^{(3)}_k]^3(-4h^9_k p^{(10)}_{k|k-1} - 144h^7_k p^{(8)}_{k|k-1} r^{(2)}_k \\
&\quad - 504h^5_k p^{(6)}_{k|k-1} r^{(4)}_k - 336h^3_k p^{(4)}_{k|k-1} r^{(6)}_k - 36h_k p^{(2)}_{k|k-1} r^{(8)}_k) \\
&\quad + k^{(1)}_k [k^{(3)}_k]^3(36h^8_k p^{(8)}_{k|k-1} r^{(2)}_k + 336h^6_k p^{(6)}_{k|k-1} r^{(4)}_k + 504h^4_k p^{(4)}_{k|k-1} r^{(6)}_k \\
&\quad + 144h^2_k p^{(2)}_{k|k-1} r^{(8)}_k + 4r^{(10)}_k)
\end{aligned} \tag{3.100}
$$

$$
\begin{aligned}
p^{(4)}_{k|k_{12}} &= [k^{(3)}_k]^4(h^{12}_k p^{(12)}_{k|k-1} + 66h^{10}_k p^{(10)}_{k|k-1} r^{(2)}_k + 495h^8_k p^{(8)}_{k|k-1} r^{(4)}_k \\
&\quad + 924h^6_k p^{(6)}_{k|k-1} r^{(6)}_k + 495h^4_k p^{(4)}_{k|k-1} r^{(8)}_k + 66h^2_k p^{(2)}_{k|k-1} r^{(10)}_k + r^{(12)}_k)
\end{aligned} \tag{3.101}
$$

3.5.2.1 Special Case - Gaussian Noise

In the case where the process noise, the measurement noise, and the initial conditions have a Gaussian distribution, the symmetrical filter equations reduce to the standard Kalman filter equations.

If the prediction error $\tilde{x}_{k|k-1}$ is Gaussian, the central moments of the error can be expressed in terms of the variance $p^{(2)}_{k|k-1}$ as

$$
E[\tilde{x}^n_{k|k-1}] = \begin{cases} (1 \times 3 \times \cdots (n-1))p^{(2)}_{k|k-1}{}^{n/2} & n \text{ even} \\ 0 & n \text{ odd} \end{cases} \tag{3.102}
$$

The expectations in (3.89) now become

$$
\begin{aligned}
E[\tilde{x}_{k|k-1}\tilde{z}_k] &= h_k p^{(2)}_{k|k-1} \\
E[\tilde{z}_k^2] &= h_k^2 p^{(2)}_{k|k-1} + r_k^{(2)} \\
E[\tilde{x}_{k|k-1}\tilde{z}_k^3] &= 3h_k^3 {p^{(2)}_{k|k-1}}^2 + 3h_k r_k^{(2)} p^{(2)}_{k|k-1} \\
&= 3E[\tilde{x}_{k|k-1}\tilde{z}_k]E[\tilde{z}_k^2] \\
E[\tilde{z}_k^4] &= 3h_k^4 {p^{(2)}_{k|k-1}}^2 + 6h_k^2 r_k^{(2)} p^{(2)}_{k|k-1} + 3{r_k^{(2)}}^2 \\
&= 3E[\tilde{z}_k^2]^2 \\
E[\tilde{z}_k^6] &= 15h_k^6 {p^{(2)}_{k|k-1}}^3 + 45h_k^4 r_k^{(2)} {p^{(2)}_{k|k-1}}^2 + 45h_k^2 {r_k^{(2)}}^2 p^{(2)}_{k|k-1} + 15{r_k^{(2)}}^2 \\
&= 15E[\tilde{z}_k^2]^3
\end{aligned}
\tag{3.103}
$$

Substituting these expressions into (3.88) the gains become

$$
\begin{aligned}
k_k^{(1)} &= E[\tilde{x}_{k|k-1}\tilde{z}_k]\ E[\tilde{z}_k^2] \\
k_k^{(3)} &= 0
\end{aligned}
\tag{3.104}
$$

$k_k^{(1)}$ is now the gain for the standard Kalman Filter. It is observed that 4^{th} and higher order moments are no longer required for the gain equations. Thus it is not necessary to propagate 4^{th} order prediction and filter moments. In addition, since $k_k^{(3)} = 0$, the 4^{th} and higher order moment terms in the 2^{nd} moment equation (3.92) vanish and this equation reduces to the filter moment equation for the standard Kalman filter.

3.6 Approximation of Prediction Moments

Since the filter equation (3.61) contains quadratic terms involving the innovations, the n^{th} order filter moment is a function of prediction moments up to

order $2n$. Similarly, since the filter equation (3.86) contains 3^{rd} order functions of the innovations, filter moments of order n are functions of prediction moments up to order $3n$. This problem cannot be solved by simply computing the high order prediction moments needed in the filter moment equations, since these high order prediction moments require the availability of filter moments of the same order.

One method to deal with this problem is to truncate the filter moment equations, including only 4^{th} and lower order prediction moments. This was done in the derivation of the vector-based non-Gaussian filter equations. Another approach is to approximate the higher order prediction moments. Using this approach 5^{th} and higher order moments are approximated as functions of the 2^{nd}, 3^{rd}, and 4^{th} order prediction moments.

Two approximations are considered. The first approximation was stated by Rao and Yar [27]. For symmetrical distributions they approximate higher than 4^{th} order moments of a random variable v_k using the relation

$$E[v_k^n] = \begin{cases} (n-1)\,E[v_k^2]\,E[v_k^{n-2}] & n \text{ even} \\ 0 & n \text{ odd} \end{cases} \tag{3.105}$$

This formula is exact if v_k has a Gaussian distribution.

Another approximation is described in [74], pp. 246-258. This expansion, called the Gram-Charlier series, approximates the density of a non-Gaussian distribution in terms of a Gaussian function, its derivatives, and the moments of the original density function.

Let x be a random variable with mean μ and variance σ and arbitrary density function $w(x)$. Let y be the normalized random variable with n^{th} moment ν_n defined

by

$$y = \frac{x - \mu}{\sigma}$$
$$\nu_n = E\left\{\left(\frac{x - \mu}{\sigma}\right)^n\right\} \tag{3.106}$$

If $\varphi(y)$ represents a Gaussian density function, the Gram-Charlier series approximation for the density function $w(y)$ is given by

$$w(y) = \varphi(y)\left(1 - a_3 H_3(y) + a_4 H_4(y) - a_5 H_5(y) + \cdots + (-1)^n H_n(y) + \cdots\right)$$

where $H_n(y)$ are the Hermite polynomials expressed as

$$H_{n+1}(y) = yH_n(y) - nH_{n-1}(y) \tag{3.107}$$

with $H_0(y) = 1$, $H_1(y) = y$. The coefficients a_n are computed from the Hermite polynomials and the density $w(y)$ using

$$a_n = \frac{(-1)^n}{n!}\int_{-\infty}^{\infty} w(y)H_n(y)dy \tag{3.108}$$

It is pointed out in [74] that when a limited number of coefficients are available specific groupings of coefficients are appropriate. In particular, the Edgeworth series involves the groupings

0, 3

0, 3, 4, 6

...

The coefficients corresponding to terms $0, 3, 4$, and 6 are

$$a_0 = 1$$

$$a_3 = -\nu_3/3!$$

$$a_4 = (\nu_4 - 3)/4!$$

$$a_6 = (\nu_6 - 15\nu_4 + 30)/6!$$

The central moments of the random variable x are then evaluated using

$$E[(x-\mu)^n] = \int_{-\infty}^{\infty} (\sigma y)^n w(y) dy \tag{3.109}$$

3.7 Experimental Evaluation of the Non-Gaussian Filters

The objective of this section is to determine the performance of the approximate filters described in this chapter. In order to measure performance it is desirable to compare these filters to optimal estimators in non-Gaussian noise. Unfortunately, optimal estimators do not exist for arbitrary probability distributions. However, an optimal estimator is available for distributions consisting of Gaussian sums. Let the measurement noise, process noise, and initial estimation error be represented as a sum of I Gaussian distributions with aggregate density function

$$p(x) = \sum_{i=1}^{I} \epsilon_i N(x - a_i, B_i) \tag{3.110}$$

where

$$\sum_{i=1}^{I} \epsilon_i = 1, \qquad \epsilon_i \geq 0 \; \forall \, i. \tag{3.111}$$

Given *a priori* densities of this form, the *a posteriori* densities of the state given the data are determined by direct application of Bayes' rule. The resulting estimator is denoted the Gaussian sum filter. The Gaussian sum filter relations are given in [10].

The operation of the Gaussian sum filter can be very computationally intensive. For L densities used to describe the initial state, N densities for the measurement noise, and M densities for the process noise; the state prediction at the first stage requires the propagation of $L * M$ separate estimates. The filtered estimate for the first stage requires $L * M * N$ estimates. In general, the number of separate prediction estimates that must be computed for the k^{th} stage is given by $L * M^k * N^{k-1}$, and the number of separate filtered estimates is $L * M^k * N^k$. The Gaussian sum representation essentially results in several Kalman filters operating in parallel. A weighted sum of these filters is used to form the a *posteriori* density. The conditional mean is formed as the convex combination of the mean values of the individual terms, or Kalman filters, of the Gaussian sum. It is important to note that the weighting function used to form this convex combination is dependent on the measurement data causing the conditional mean to be a nonlinear function of the data. In contrast to the linear Kalman filter, the conditional variance of the Gaussian sum filter is a function of the measurement data. Thus the conditional variance is not expected to converge smoothly as it does with the linear Kalman filter. Additionally, it cannot be computed off line as can be done with the Kalman filter for linear systems.

The primary advantage of the Gaussian sum approach is that it forms an explicit representation of the a *posteriori* density. This representation is optimal if the errors are truly made up of a sum of Gaussian distributions. The major disadvantage of the Gaussian sum filter is the geometric progression of the number of separate Kalman filters that are required to implement the estimator. However, the number of filters can be limited by disregarding those terms in the Gaussian sum that have very small weights.

Our interest in experimental evaluation of the approximate filters is to determine what degree of improvement these filters offer over the standard Kalman filter and to determine how close these approximate filters match the performance of the optimal estimator using Gaussian sums. Also of interest is to determine if the truncated forms of the asymmetrical and symmetrical filter approximations given in Sections 3.2 and 3.3 give similar performance to the nontruncated expressions of Sections 3.5.1 and 3.5.2.

A scalar model is used to evaluate the performance of the filters. The plant and measurement equations for this model have the form

$$\begin{aligned} x_k &= 0.5x_{k-1} + w_{k-1} \\ z_k &= x_k + v_k \end{aligned} \tag{3.112}$$

where w_k and v_k are mutually independent, zero mean possibly non-Gaussian random processes, and the initial estimation error for x_0 may also be non-Gaussian. The non-Gaussian distributions are modeled as the sum of two Gaussian distributions with unit variance. In general the non-Gaussian distribution for a random variable y is given by

$$p(y) = \sum_{i=1}^{I} \epsilon_i N(\mu_i, 1) \tag{3.113}$$

where $\sum_{i=1}^{I} \epsilon_i = 1$. For the special case of two distributions ($I = 2$) the parameter D is defined as the separation between the means of the distributions. In this case for $p(y)$ to have zero mean $\mu_1 = -\epsilon_2 * D$, and $\mu_2 = \epsilon_1 * D$. If $\epsilon_1 = \epsilon_2$, $p(y)$ is symmetric. If $D = 0$, $p(y)$ is Gaussian.

In general the density $p(y)$ has zero mean and the next three moments are

$$E[y^2] = \sum_{i=1}^{I} \epsilon_i(1 + \mu_i^2)$$

$$E[y^3] = \sum_{i=1}^{I} \epsilon_i(\mu_i^3 + 3\mu_i) \tag{3.114}$$

$$E[y^4] = \sum_{i=1}^{I} \epsilon_i(3 + 6\mu_i^2 + \mu_i^4)$$

3.7.1 Asymmetrical Filter Results

An asymmetrical distribution with $\epsilon_1 = 0.2$, $\epsilon_2 = 0.8$, and $I = 2$ in equation (3.113). The system represented by equations (3.112) is used to evaluate the different filters for various combinations of non-Gaussian process noise v_k, measurement noise w_k, and initial estimation error $\tilde{x}_0$. The three noise models are given in Table 3.1 below.

Table 3.1. Noise Models for Non-Gaussian Filter Evaluation

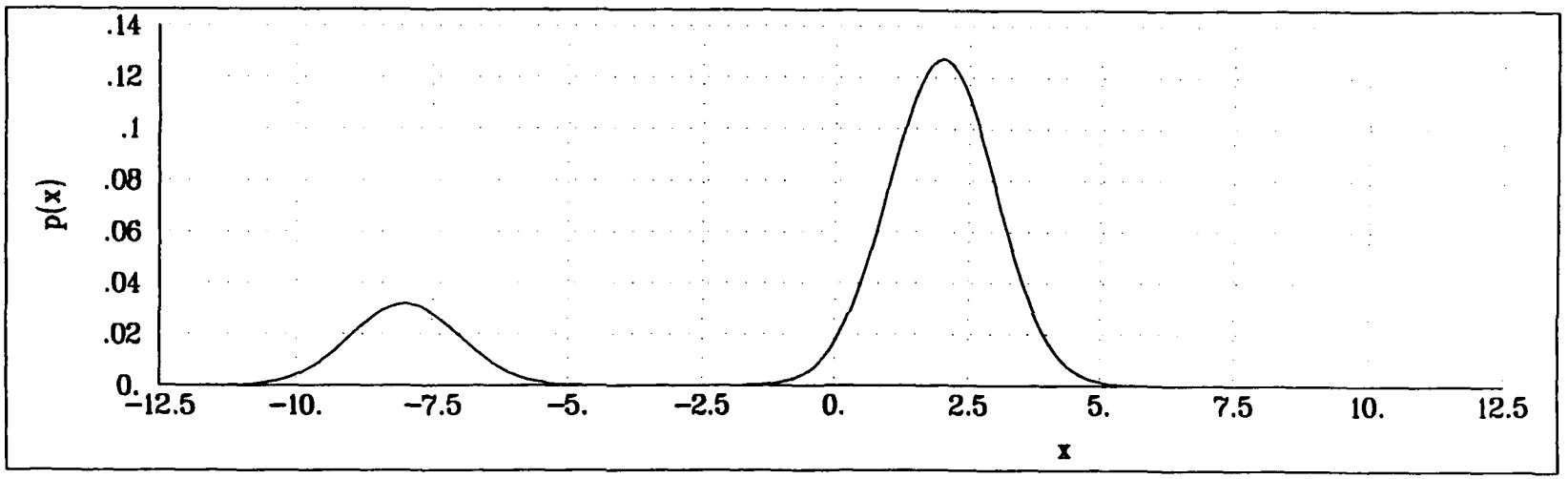
Model	v_k	w_k	$\tilde{x}_0$
1	$\sum_{i=1}^{I} \epsilon_i N(\mu_i, 1)$	$N(0,1)$	$N(\hat{x}_0, 1)$
2	$N(0,1)$	$\sum_{i=1}^{I} \epsilon_i N(\mu_i, 1)$	$N(\hat{x}_0, 1)$
3	$N(0,1)$	$N(0,1)$	$\sum_{i=1}^{I} \epsilon_i N(\mu_i + \hat{x}_0, 1)$

The asymmetrical filter of Section 3.5.1 is evaluated in three different configurations. In the first configuration, denoted Asym/T1, the asymmetrical filter equations of Section 3.5.1 are modified so that the 3^{rd} order filter moment contains only functions of 3^{rd} order prediction moments and 3^{rd} order measurement error, and the 4^{th} order filter moment contains only functions of 4^{th} order prediction mo-

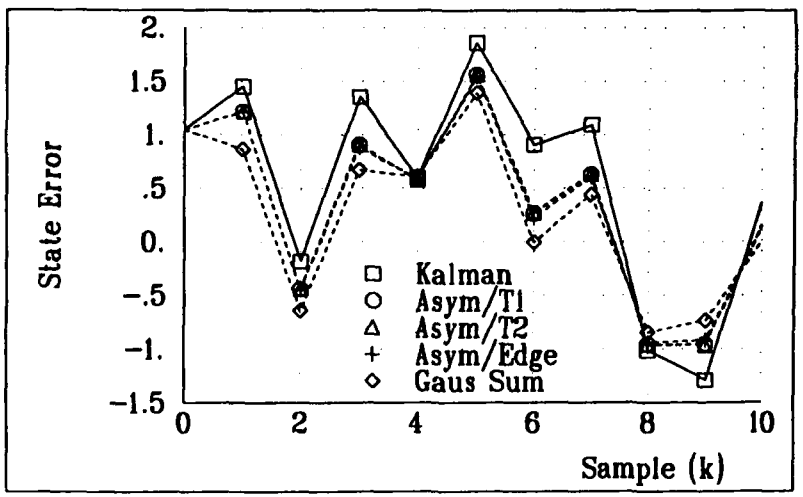
ments and 4^{th} order measurement error. Thus, the terms $p^{(3)}_{k|k_4}, p^{(3)}_{k|k_5}$, and $p^{(3)}_{k|k_6}$ are set to zero in equation (3.73), and $p^{(4)}_{k|k_5}, p^{(4)}_{k|k_6}, p^{(4)}_{k|k_7}$, and $p^{(4)}_{k|k_8}$ are set to zero in equation (3.78). The second asymmetrical filter configuration, denoted Asym/T2, is similar to Asym/T1 with the exception that the 4^{th} moment term $p^{(3)}_{k|k_4}$ is retained (3.73). The vector formulation for this truncated filter configuration is developed in Section 3.2. The last asymmetrical filter configuration, denoted Asym/Edge, uses the complete filter configuration of Section 3.5.1 with 5^{th} and higher order moments being approximated by the Edgeworth expansion using coefficients a_0 and a_3. The second Edgeworth expansion, which uses the terms a_0, a_3, a_4, and a_6, cannot be used because the a_6 coefficient requires the availability of the 6^{th} order moment. These three asymmetrical filters are compared to the performance of the standard Kalman filter and to the Gaussian sum filter.

Figure 3.1 displays the non-Gaussian noise distribution, the state estimation error $\tilde{x}_{k|k}$, and the filter variance $p^{(2)}_{k|k}$ for a typical simulation of the system in equation (3.112) for Model 1. The separation between the distributions for the non-Gaussian noise was $D = 10$. Figure 3.1 illustrates that the asymmetrical filter performance is significantly better than the standard Kalman filter, but not as good as the Gaussian sum filter. The three asymmetrical filter configurations perform about equally. It is observed that since Model 1 is indeed a Gaussian sum model, the Gaussian sum filter produces an optimal estimate. The primary purpose of this test is to compare the standard Kalman filter to the HOF, since both filters use only the error moments and not the density function to perform estimation. The Gaussian sum filter performance is included only as a reference.

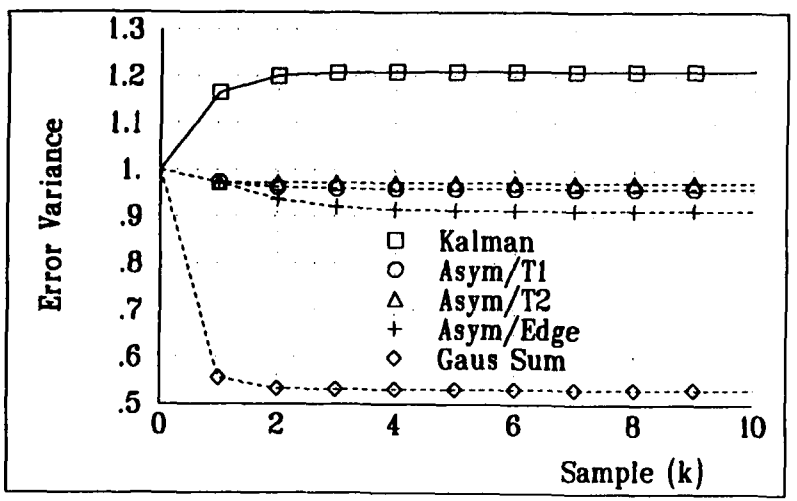
A Monte-Carlo analysis was performed to determine the sample variance of the estimation error. Fifty separate simulation runs of the system were made for



(a) Measurement Noise Probability Density Function



(b) Estimation Error $\tilde{x}_{k|k}$



(c) Error Variance $P_{k|k}$

Figure 3.1 Typical Asymmetrical Filter Results, Model 1, $D = 10$, Bimodal Measurement Noise

each model. The estimation error was accumulated over the last five samples of each run resulting in a total of 250 samples of the estimation error. The sample variances of the filter error $\hat{P}^{(2)}_{k|k}$ and the prediction error $\hat{P}^{(2)}_{k|k-1}$ are presented in Table 3.2 for every estimator for each model.

Table 3.2. Sample Variances for the Asymmetrical Filter - $D = 10$

Filter	Model 1		Model 2		Model 3	
Type	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$
Kalman	1.25	1.35	0.983	18.4	0.575	1.20
Asym/T1	0.952	1.29	0.828	18.5	0.575	1.20
Asym/T2	0.954	1.29	0.828	18.5	0.575	1.20
Asym/Edge	0.944	1.29	0.828	18.5	0.575	1.20
Gaus Sum	0.575	1.20	0.575	18.5	0.575	1.20

The Monte Carlo results are consistent with the observations made from the single run results in that the asymmetrical filter performed very well in relation to the standard Kalman filter. The fact that the Asym/Edge model gives the same performance as any of the truncated forms suggests that the truncated filters are sufficient to characterize the asymmetrical filter. Although there may be other approximations for higher moments that give better results than the Edgeworth series, this suggests that there is no need to go through the lengthy and cumbersome vector expansion for 5^{th} and higher order components for the 3^{rd} and 4^{th} filter moments.

A similar study was done on the asymmetrical filters for distribution $D = 5$. The Monte Carlo results for this configuration are given in Table 3.3.

Table 3.3 Sample Variances for the Asymmetrical Filter - $D = 5$

Filter Type	Model 1		Model 2		Model 3	
	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$
Kalman	1.08	1.34	0.910	5.50	0.575	1.20
Asym/T1	0.923	1.32	0.823	5.49	0.575	1.20
Asym/T2	0.923	1.32	0.823	5.49	0.575	1.20
Asym/Edge	0.922	1.32	0.823	5.49	0.575	1.20
Gaus Sum	0.742	1.28	0.752	5.49	0.575	1.20

As expected the sample variances of the estimation error are closer together than they were in Table 3.2 where $D = 10$. As the distribution separation D approaches 0 all of the results would be the same and all of the estimators would be optimal.

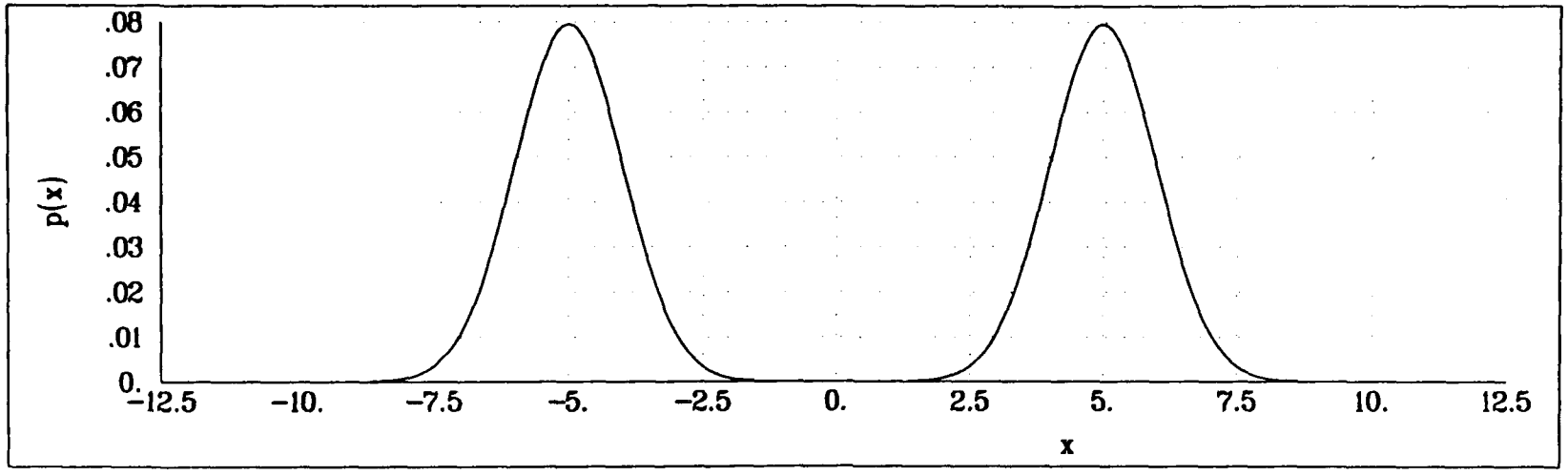
3.7.2 Symmetrical Filter Results

A symmetrical non-Gaussian distribution is generated with parameter values $\epsilon_1 = 0.5$, $\epsilon_2 = 0.5$, and $I = 2$ in equation (3.112). The system represented by equation (3.113) was evaluated for various combinations of non-Gaussian process noise v_k, measurement noise w_k, and initial estimation error $\tilde{x}_0$ as expressed in Table 3.1.

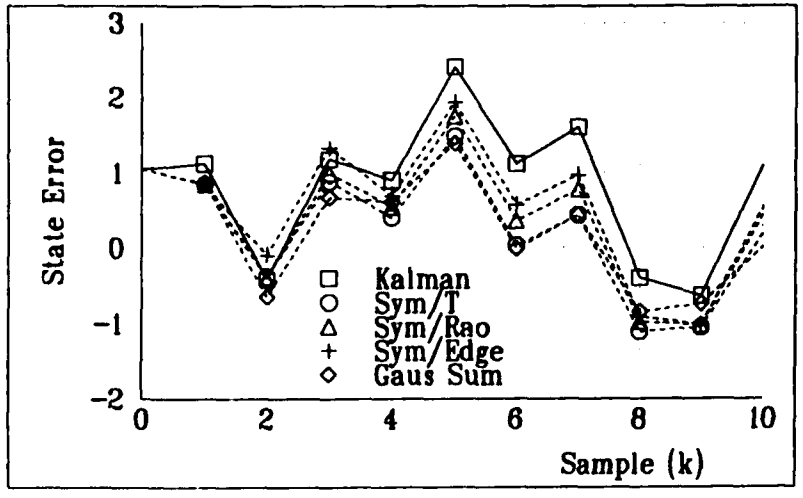
The symmetrical filter was evaluated in three different configurations. In the first configuration, denoted Sym/T, the symmetrical filter equations of Section 3.5.2 were modified so that the second order filter moment contains only functions of 2^{nd} and 4^{th} order prediction moments and measurement moments, and the 4^{th} order filter moment contains only functions of 4^{th} order functions of the prediction

and measurement errors. Thus terms $p^{(2)}_{k|k_4}$, and $p^{(2)}_{k|k_6}$ were set to zero in equation (3.92), and $p^{(4)}_{k|k_6}, p^{(4)}_{k|k_8}, p^{(4)}_{k|k_{10}}$, and $p^{(4)}_{k|k_{12}}$ were set to zero in equation (3.96). The second symmetrical filter configuration, denoted Sym/Rao used the complete filter configuration of Section 3.5.2. Moments of order higher that 4^{th} were approximated using the formula described in equation (3.105), which was obtained from Rao and Yar [27]. The last symmetrical filter configuration, denoted Sym/Edge, used the complete filter configuration of Section 3.5.2 with moments of order higher than 4^{th} were approximated by the Edgeworth expansion using coefficients a_0 and a_3. It is noted that $E[\tilde{z}^6_k]$ in equation (3.89) requires the availability of the 6^{th} moment of the prediction error, and the 6^{th} moment of the measurement error. For the truncated model Sym/T the Rao approximation was used for these moments.

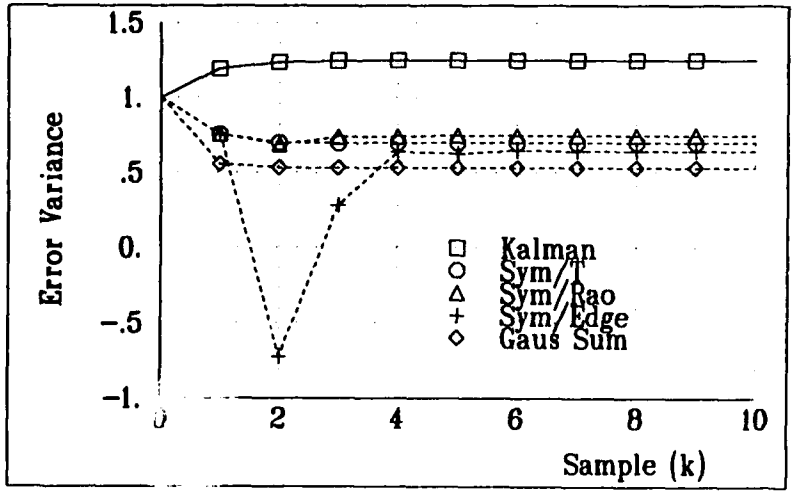
The Monte Carlo results for 250 samples are given in Table 3.4. This table shows that the symmetrical HOF performs better than the standard Kalman filter, but not quite as good as the optimal Gaussian sum filter. Among the three asymmetrical filter configurations, the truncated form and the form that uses the Rao approximation perform the same. However, the Sym/Edge filter performance is somewhat poorer than the other non-Gaussian filters. This discrepancy is probably due to the fact that the Edgeworth expansion is based on a Gaussian kernel and this approximation degrades as the separation D increases. A similar study was performed for symmetrical distributions with $D = 5$. The Monte Carlo results are given in Table 3.5.



(a) Measurement Noise Probability Density Function



(b) Estimation Error $\tilde{x}_{k|k}$



(c) Error Variance $P_{k|k}$

Figure 3.2 Typical 6^{th} Order Symmetrical Filter Results, Model 1, $D = 10$, Bimodal Measurement Noise

Table 3.4. Sample Variances for the Symmetrical Filter - $D = 10$

Filter	Model 1		Model 2		Model 3	
Type	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$
Kalman	1.33	1.35	0.947	25.7	0.575	1.20
Sym/T	1.22	1.32	0.908	25.7	0.575	1.20
Sym/Rao	1.22	1.32	0.908	25.7	0.575	1.20
Sym/Edge	1.29	1.34	0.933	25.7	0.575	1.20
Gaus Sum	0.575	1.20	0.575	25.6	0.575	1.20

Table 3.5. Sample Variances for the Symmetrical Filter - $D = 5$

Filter	Model 1		Model 2		Model 3	
Type	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$	$\hat{P}^{(2)}_{k\|k}$	$\hat{P}^{(2)}_{k\|k-1}$
Kalman	1.20	1.35	0.876	7.20	0.575	1.20
Sym/T	1.10	1.32	0.838	7.21	0.575	1.20
Sym/Rao	1.10	1.32	0.838	7.21	0.575	1.20
Sym/Edge	1.15	1.33	0.856	7.20	0.575	1.20
Gaus Sum	0.914	1.26	0.782	7.21	0.575	1.20

As with the asymmetrical filter, as the separation D becomes smaller, the sample variances of the estimation error are clustered closer together for all models since as D approaches 0 the mixture distribution becomes "more" Gaussian. For this case the sample variance for the Sym/Edge model is now about halfway between the sample variance for the standard Kalman filter and the Sym/T model, whereas the sample variance for Sym/Edge for $D = 10$ was closer to the Kalman sample variance. This is expected since the non-Gaussian distribution is "more" Gaussian

as D becomes smaller and thus the Edgeworth approximation is more valid.

Better results may be obtained for the symmetrical filter by propagating 6^{th} order filter and prediction moments. The 6^{th} order prediction moment can be obtained directly from the prediction error equation $\tilde{x}_{k|k-1} = \phi_{k-1}\tilde{x}_{k-1|k-1} + w_{k-1}$. The 6^{th} moment is given by

$$p^{(6)}_{k|k-1} = \phi^6_{k-1}p^{(6)}_{k-1|k-1} + 15\phi^4_{k-1}p^{(4)}_{k-1|k-1}q^{(2)}_{k-1} + 15\phi^2_{k-1}p^{(2)}_{k-1|k-1}q^{(4)}_{k-1} + q^{(6)}_{k-1} \quad (3.115)$$

If the 3^{rd} power of the innovations in equation (3.87) are disregarded, the 6^{th} order filter moment can be expressed in terms of only 6^{th} order functions of the prediction and measurement errors. This truncation equation lead to

$$p^{(6)}_{k|k} = a^6_k p^{(6)}_{k|k-1} + 15a^4_k k^{(1)\,2}_k p^{(4)}_{k|k-1}r^{(2)}_k + 15a^2_k k^{(1)\,4}_k p^{(2)}_{k|k-1}r^{(4)}_k + k^{(1)\,6}_k r^{(6)}_k \quad (3.116)$$

where $a_k = 1 - k^{(1)}_k h_k$.

In addition the 6^{th} order terms $p^{(2)}_{k|k_6}$ are retained in (3.92), and $p^{(4)}_{k|k_6}$ in (3.96).

With this configuration the 6^{th} order symmetrical filters was tested for $D = 5, 10$, and 15 for a bimodal measurement and process noise distributions. An example simulation run for bimodal measurement noise (Model 1) is shown in Figure 3.2. Table 3.6 compares the results of the linear Kalman filter, the 4^{th} and 6^{th} order symmetrical filters, and the Gaussian sum filter for Model 1. Included in this figure are the values of the bimodal noise variance $E[v^2_k]$. Define a coefficient of excess Γ_x for a random variable x as

$$\Gamma_x = \frac{3E[x^2]}{E[x^4]} \quad (3.117)$$

Γ_x represents the deviation of the actual fourth moment from the Gaussian fourth moment. This value is also included in the Table 3.6.

Table 3.6. Comparison of Sample Filter Variances for 4^{th} and 6^{th} Order Filters

Bimodal Measurement Noise Distribution

D	$E[v_k^2]$	Γ_{v_k}	Filter Order	Sample Filter Variances				
				Kalman	Sym/T	Sym/Rao	Sym/Edge	Gaus Sum
5	7.25	1.98	4^{th}	1.20	1.10	1.10	1.15	0.914
			6^{th}		1.01	1.01	1.01	
10	26	2.61	4^{th}	1.33	1.22	1.22	1.29	0.575
			6^{th}		0.773	0.824	0.897	
15	101	2.89	4^{th}	1.34	1.28	1.28	1.32	0.575
			6^{th}		0.673	0.773	0.880	

Table 3.6 shows that as the separation D increases the 6^{th} order filters perform significantly better than the 4^{th} order filters. The truncated filter, 'Sym/T', gives the best overall performance. In fact, for $D = 15$ the sample variance of 'Sym/T' (0.673) is very close to that of the Gaussian sum filter (0.575). It is observed that as D increases from 5 to 10, Γ_{v_k} increases from 1.98 to 2.89. Thus, as the bimodal distribution becomes "more" non-Gaussian, the 6^{th} order filter performs much better relative to the linear Kalman filter. It is important to note that the high order filters may become unstable under certain conditions. This is evident in Figure 3.2(c), in which the Sym/Edge configuration behaves erratically. However, unlike the Gaussian sum filter, the error variance $P_{k|k}$ for the non-Gaussian filters can be evaluated off-line without any measurement data if the system is linear, so stablity can be determined before the filter is implemented on actual data.

Table 3.7 gives the results for Model 2, bimodal process noise. The results are consistent with those in Table 3.6.

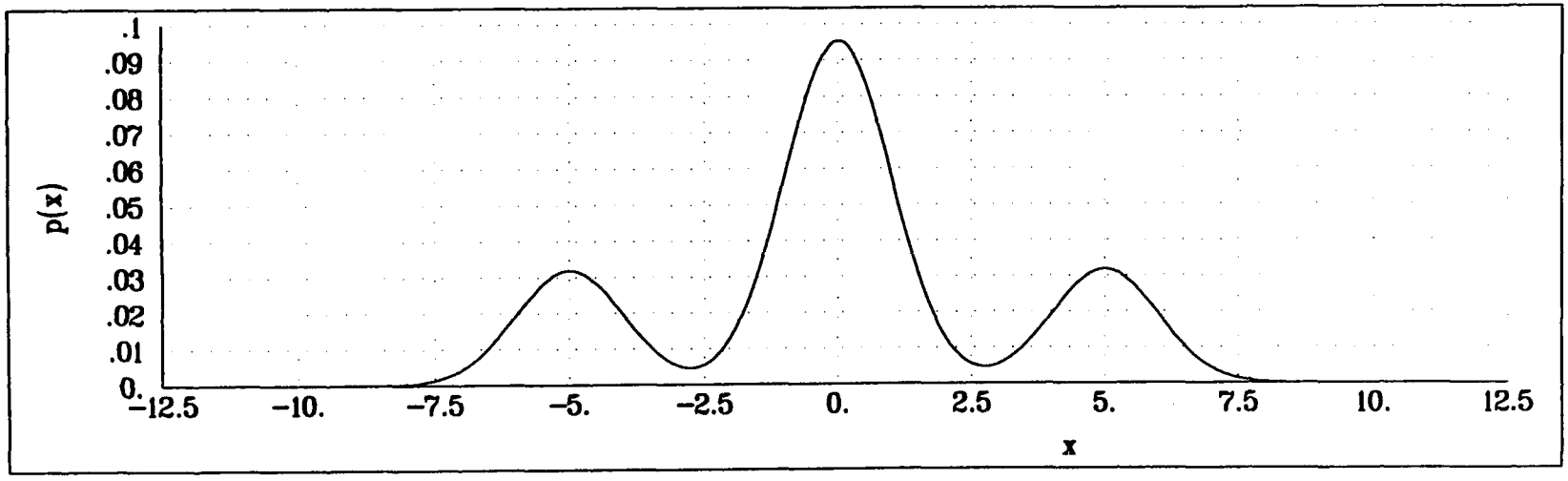
Table 3.7. Comparison of Sample Filter Variances for 4^{th} and 6^{th} Order Filters

Bimodal Process Noise Distribution

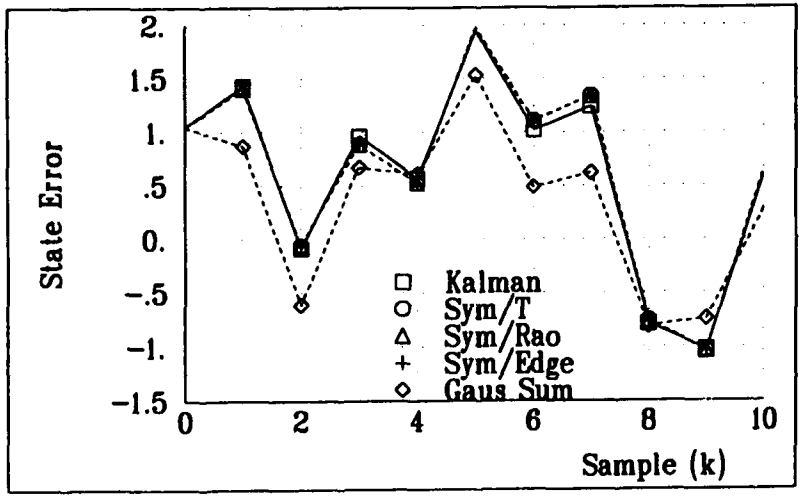
D	$E[w_k^2]$	Γ_{w_k}	Filter Order	Sample Filter Variances				
				Kalman	Sym/T	Sym/Rao	Sym/Edge	Gaus Sum
5	7.25	1.98	4^{th}	0.876	0.837	0.837	0.856	0.782
			6^{th}		0.789	0.791	0.794	
10	26	2.61	4^{th}	0.947	0.908	0.908	0.933	0.575
			6^{th}		0.684	0.718	0.798	
15	101	2.89	4^{th}	0.964	0.939	0.939	0.956	0.575
			6^{th}		0.643	0.739	0.790	

A commonly encountered non-Gaussian distribution is the so-called heavy-tailed Gaussian distribution. This distribution is composed of a large central lobe and two smaller lobes separated by an equal distance on each side of the main lobe. To generate this distribution $I = 3$, $\epsilon_1 = \epsilon_3 = 0.2, \epsilon_2 = 0.6, \mu_1 = -D/2, \mu_2 = 0, \mu_3 = D/2$ were used in equation (3.113). Figure 3.3 displays the non-Gaussian noise distribution, the estimation error $\tilde{x}_{k|k}$, and the filter variance $p_{k|k}^{(2)}$ for a typical simulation of the system in equation (3.112) for model 1. The separation between the distributions for the non-Gaussian noise was $D = 10$. Tables 3.8 and 3.9 compare the performance of the non-Gaussian filters to the Kalman and Gaussian sum filters for Models 1 and 2 respectively.

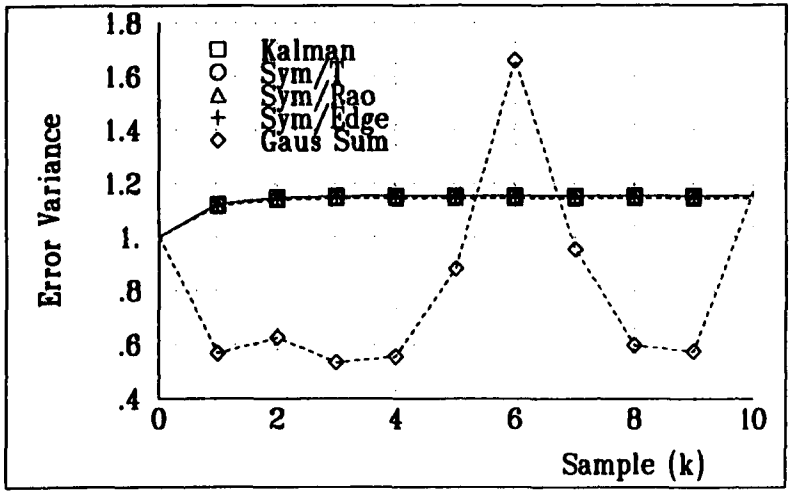
Tables 3.8 and 3.9 show that the Gaussian sum filters do not offer any significant improvement over the Kalman filter for the heavy-tailed distributions used for these examples. For the heavy-tailed distributions Γ_{v_k} and Γ_{w_k} are approximately equal to 1. This again demonstrates that the relative performance of the



(a) Measurement Noise Probability Density Function



(b) Estimation Error $\tilde{x}_{k|k}$



(c) Error Variance $P_{k|k}$

Figure 3.3 Typical 6^{th} Order Symmetrical Filter Results, Model 1, $D = 10$, Heavy-Tailed Measurement Noise

Table 3.8. Comparison of Sample Filter Variances for 4^{th} and 6^{th} Order Filters

Heavy-Tailed Measurement Noise Distribution

D	$E[v_k^2]$	Γ_{v_k}	Filter Order	Sample Filter Variances				
				Kalman	Sym/T	Sym/Rao	Sym/Edge	Gaus Sum
5	3.5	1.10	4^{th}	0.964	0.966	0.967	0.966	0.989
			6^{th}		0.967	0.967	0.967	
10	11	1.16	4^{th}	1.20	1.20	1.20	1.20	0.918
			6^{th}		1.21	1.21	1.21	
15	23.5	1.18	4^{th}	1.27	1.27	1.27	1.27	0.603
			6^{th}		1.28	1.28	1.28	

Table 3.9. Comparison of Sample Filter Variances for 4^{th} and 6^{th} Order Filters

Heavy-Tailed Process Noise Distribution

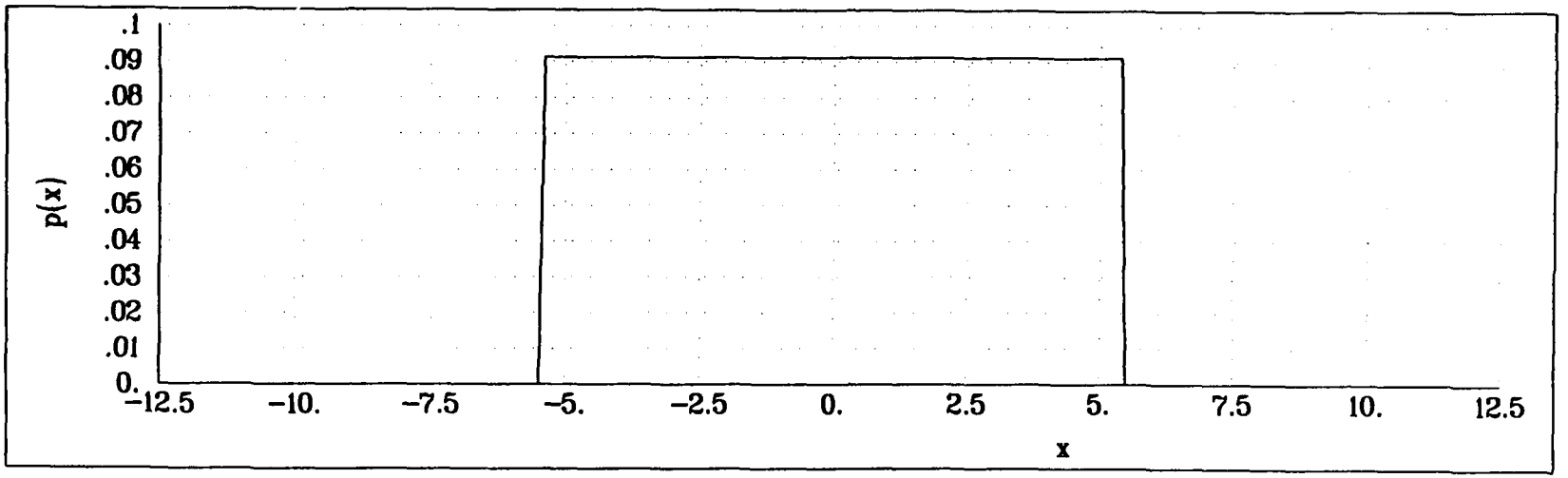
D	$E[w_k^2]$	Γ_{w_k}	Filter Order	Sample Filter Variances				
				Kalman	Sym/T	Sym/Rao	Sym/Edge	Gaus Sum
5	3.5	1.10	4^{th}	0.853	0.854	0.854	0.854	0.849
			6^{th}		0.854	0.854	0.854	
10	11	1.16	4^{th}	0.957	0.955	0.955	0.956	0.849
			6^{th}		0.954	0.954	0.954	
15	23.5	1.18	4^{th}	0.980	0.978	0.978	0.978	0.639
			6^{th}		0.972	0.972	0.972	

non-Gaussian filters can be assessed by comparing the fourth moment to that of a Gaussian distribution. If the coefficient of excess is close to 1, then the standard Kalman filter gives similar performance to the symmetrical high order filter.

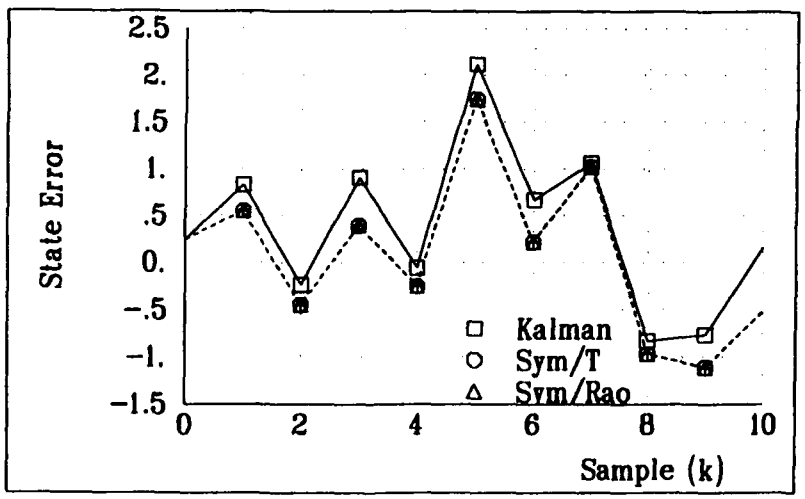
The last non-Gaussian distribution considered is the uniform distribution. Figure 3.4 displays the noise distribution, the estimation error $\tilde{x}_{k|k}$, and the filter variance $p^{(2)}_{k|k}$ for a typical simulation of the system in equation (3.112) for uniform measurement noise with variance equal to 10. Tables 3.10 and 3.11 compare the performance of the non-Gaussian filters to the Kalman and filters for uniformly distributed measurement noise (Gaussian process noise and initial estimation error with unit variance), and uniformly distributed process noise (Gaussian measurement noise and initial estimation error with unit variance), respectively.

Table 3.10. Comparison of Sample Filter Variances for 4^{th} and 6^{th} Order Filters Uniform Measurement Noise Distribution

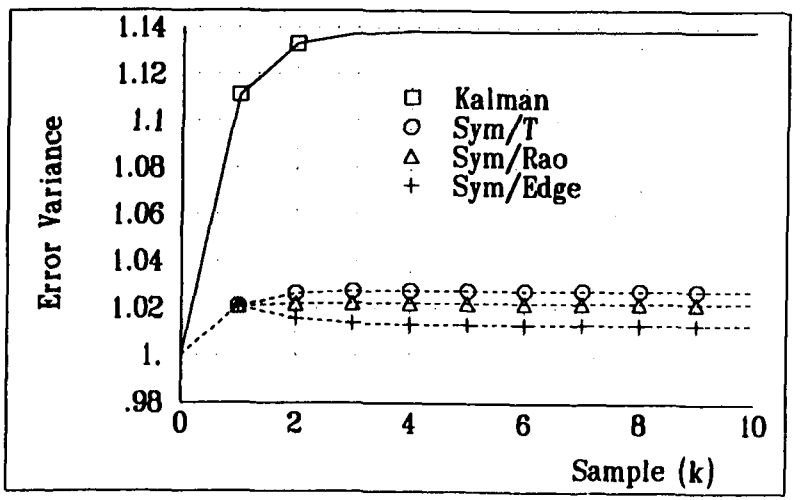
$E[v_k^2]$	Γ_{v_k}	Filter Order	Sample Filter Variances			
			Kalman	Sym/T	Sym/Rao	Sym/Edge
5	1.67	4^{th}	0.944	0.868	0.868	0.893
		6^{th}		0.825	0.825	0.824
10	1.67	4^{th}	1.04	0.984	0.984	1.01
		6^{th}		0.912	0.912	0.911
15	1.67	4^{th}	1.08	1.03	1.03	1.06
		6^{th}		0.963	0.963	0.962



(a) Measurement Noise Probability Density Function



(b) Estimation Error $\tilde{x}_{k|k}$



(c) Error Variance $P_{k|k}$

Figure 3.4 Typical 6^{th} Order Symmetrical Filter Results, Model 1, $D = 10$, Uniform Measurement Noise

Table 3.11. Comparison of Sample Filter Variances for 4^{th} and 6^{th} Order Filters Uniform Process Noise Distribution

$E[w_k^2]$	Γ_{w_k}	Filter Order	Sample Filter Variances			
			Kalman	Sym/T	Sym/Rao	Sym/Edge
5	1.67	4^{th}	0.866	0.838	0.838	0.848
		6^{th}		0.815	0.815	0.816
10	1.67	4^{th}	0.932	0.910	0.910	0.919
		6^{th}		0.876	0.876	0.876
15	1.67	4^{th}	0.955	0.937	0.937	0.945
		6^{th}		0.902	0.902	0.903

The non-Gaussian filters give better performance than the Kalman filter for uniformly distributed noise. Again the relative performance is related to the difference between the fourth moment $E[v_k^4]$ and the fourth moment of the Gaussian distribution $3E[v_k^2]^2$.

3.8 Conclusion

Two approximate methods for filtering have been presented for estimation in the presence of asymmetric and symmetric distributions of non-Gaussian noise. Simulation studies have shown that the HOFs can perform very well for estimation in non-Gaussian noise. The real utility of the filters developed in this chapter comes when either the noise cannot be adequately represented as Gaussian sums, or when only the moments of the noise are known, and not the actual density functions. Although these filters are more complicated to implement than the standard Kalman filter, they are not nearly as computationally intensive as the Gaussian sum filter

in which the number of parallel filters grows geometrically as the number of stages increase.

An obvious method to improve the performance of the non-Gaussian filters developed here is to use higher powers of the innovations in developing the filter equations. That is, let I be greater than 3 in

$$\tilde{\mathbf{x}}_{k|k} = \tilde{\mathbf{x}}_{k|k-1} - \sum_{i=0}^{I} K_k^{(i)} \tilde{\mathbf{z}}_k^{\otimes i}.$$

However, this would make the vector derivation of the filter equations very unwieldy. In addition this derivation would require the availability of still higher order terms in the solution of the filter variance equations. That is for $I = 4$ the filter variance equation would require up to 8^{th} order prediction moments. The 4^{th} order moment of the variance would require up to 16^{th} order prediction moments. In general, when $I > 1$ it is necessary to either truncate the expressions for the filter moments so that only those powers of prediction and measurement error moments are included for which similar powers of the filter moments exist, or the higher powers of the prediction and measurement error moments must be approximated.

For non-Gaussian distributions made up of known Gaussian sums, the non-Gaussian filters presented here give a reasonable compromise between the optimal but very computationally intensive Gaussian sum filter, and the suboptimal but easily implemented standard Kalman filter. In addition, when only the moments of the distributions are known and a Gaussian sum filter cannot be used, the non-Gaussian filters offer a means to obtain improved performance over the standard Kalman filter.

Chapter 4

Nonlinear Filtering Methods for Harmonic Retrieval

This chapter addresses the problem of high resolution parameter estimation of superimposed sinusoids using nonlinear filtering techniques. Six separate nonlinear filters are evaluated for the estimation of the parameters of sinusoids in white and colored Gaussian noise. Experimental evaluation demonstrates that the nonlinear filters perform very satisfactorily (close to the Cramer-Rao bound) for reasonable values of the initial estimation error. A major advantage of using nonlinear filtering methods for harmonic retrieval is that the filters can be applied to time varying process models as well.

Some of the more recent work done in parametric methods for harmonic superesolution includes modified singular value decomrpostion techniques [31], and cumulant based techniques [32]. Generally these approaches perform well at high SNR's, with close correspondence to the CR bound. However, the performance is severely degraded at low SNR's.

Solution of the harmonic retrieval problem is approached using 3 nonlinear filters including the extended Kalman filter, the Gaussian second order filter, and the minimum variance filter. Three iterated forms of the extended Kalman filter are also applied to this problem. The main advantage of using recursive filtering techniques over more traditional batch-type estimators is that time varying system parameters can be modeled. The nonlinear filters are applied to data consisting of two exponentially damped sinusoids in white noise. The results are compared to the Cramer-Rao (CR) bound and to results obtained by other authors using singular

value decomposition (SVD) techniques. The performance of the nonlinear estimators is also evaluated in colored noise with known and unknown noise filter coefficients. In addition, a technique is presented to perform estimation when the noise statistics are unknown. In this case the noise statistics are estimated along with the state estimates. Overall it is found that the nonlinear filters give performance very close to the CR bound whenever the initial state covariance is small. The techniques are found to be very effective in colored noise with known and unknown coefficients, and when the noise statistics are unknown.

The problem of high resolution frequency estimation has received a considerable amount of attention in recent years. The classical method for frequency estimation is based on nonparametric periodogram estimates and their variations [31]. The frequency estimates are formed from the power spectral density estimates obtained from the Fourier transform of the data or from the Fourier transform of the autocorrelation sequence. The frequency resolution of these techniques is directly related to the number of data samples in the received time series, so the periodogram method is generally not considered a high resolution frequency estimator. In a related technique developed by Capon [32], called maximum likelihood (ML) spectral estimation, the PSD is estimated by effectively measuring the power out of a set of narrowband filters. Foias *et al.* [33] have demonstrated that the ML estimate converges monotonically to the point power spectrum associated with the sinusoids as the number of correlation lags approaches infinity. They further show how this convergence property can be used to determine whether a strong spectral peak corresponds to a sinusoid. A complete review of all of the basic spectrum estimation techniques is presented by Kay and Marple [34].

Parametric methods attempt to fit some assumed model to the data. Using

parametric methods, the problem becomes one of choosing a proper model and estimating the parameters of the assumed model. Most of these models can be classified into autoregressive (AR) or autoregressive moving average (ARMA) models. Some of the early work on these models was done by Ulrych and Clayton [35] for the (AR) model, and Cadzow [36] for the ARMA model. Ulrych and Clayton use a modified covariance technique in which the sum of the squared forward and backward residuals is minimized. Lang and McClellan [37] have shown that the variance of the spectral estimate obtained using this approach is smaller than that of the covariance method. Citron *et al.* [38] compare Ulrych and Clayton's method with Cadzow's method and find that Ulrych and Clayton's technique appears to perform better. Tufts and Kumaresan [39 – 43] and Hua [44] have enhanced the performance of AR models by using singular value decomposition (SVD) and a reduced rank approximation. They show that this technique has close correspondence to the CR bound at high SNR's. However, Bresler and Macovski [45] point out that the performance of most modern high resolution spectral estimation methods is severely degraded at low SNR's and/or short data lengths. They postulate that this is due to the fact that these techniques are heuristic least squares modifications of algorithms that yield exact results when there is no noise or when the available data is infinite (known covariance case). Other methods include adaptive notch filtering [46], adaptive line enhancement [47], and pencil of function methods [48 – 50].

A relatively new method used for harmonic retrieval involves the use of higher order statistics [51,52]. For non-Gaussian signal components and Gaussian noise the third order moment and fourth order cumulant of the measurements will theoretically only contain signal components and accurate estimates for the signal frequency components can be obtained. Papadopoulos and Nikias [52] show that by using these

methods they can match the performance of the Kumaresan and Tufts methods in white Gaussian noise and perform better in colored Gaussian noise. However, the high order statistics methods show severely degraded performance at low SNR's. Arun and Aung [57] have proposed a SVD based approach for tracking the parameters of sinusoids with time varying parameters.

Other algorithms have been shown to give good performance at low SNR's. Chan *et al.* [53] developed recursive expressions for the estimation of m sinusoids in white noise using $2m$ coefficients of an ARMA model. Friedlander [54] developed a recursive algorithm for maximum likelihood ARMA spectral estimation. The iterative inverse filtering method [55] is shown to produce accurate estimates of unknown frequencies at low SNR's and a small number of points.

Stankovic *et al.* [56] uses the extended Kalman filter for the estimation of the frequencies of sinusoids in white and colored noise. They use an ARMA model for the signal and the noise. Thus, they need to estimate 2 variables for each frequency. For good initial parameter estimates, the EKF method outperforms the maximum likelihood method.

In this chapter nonlinear filtering techniques are employed to estimate the parameters of exponentially damped sinusoids in colored noise. A direct model is used. This model requires one state variable for each parameter to be estimated. That is, the state variables are the frequencies, phases, damping coefficients, and amplitudes of the sinusoids. Using this model time varying characteristics of the state can be explicitly evaluated. A comparison is made among the EKF, the iterated filters, and the minimum variance filter for this problem. Various filters are evaluated in order to determine the estimator that is least susceptible to the impact

of the initial state covariance on the performance of the algorithms. The results are compared to the CR bound and to the SVD methods of Kumaresan and Tufts.

This chapter is organized as follows: Section 4.1 describes the general system model used for all of the filters for estimating time varying amplitudes, frequencies, damping coefficients, and phases for a known number of sinusoids in additive white Gaussian noise. Section 4.2 presents the nonlinear filtering equations for the general system model and the details of the implementation for the specific model involving the two exponentially damped sinusoids in white noise. This section discusses the extended Kalman filter, the Gaussian second order filter, the minimum variance filter, and three iterated forms of the extended Kalman filter. Experimental results obtained from Monte Carlo simulations demonstrate the performance of these filters. Section 4.3 presents the extended Kalman filter expressions and simulation results for harmonic retrieval in colored noise with known and unknown noise filter coefficients. A technique for estimation of the measurement covariance is described and experimentally analyzed in section 4.4.

4.1 General System Model

Consider the problem of estimating the parameters of P sinusoids from K measurements. The complex scalar measurement model is given by

$$z_k = \sum_{p=1}^{P} c_{k_p} \exp\left(-\alpha_{k_p} k + j(\omega_{k_p} k + \theta_{k_p})\right) + v_k \tag{4.1}$$

for $k = 0, 1, \cdots, K-1$. v_k is assumed to be complex white Gaussian noise with mutually independent real and imaginary components each with variance σ^2. It is assumed that the frequencies ω_{k_p} are normalized so that the effective sampling interval is one second. The 2-element measurement model $\mathbf{z}_k$ from (4.1) can be

written as

$$\begin{aligned}\mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k \\ &= \begin{bmatrix} \sum_{p=1}^{P} c_{k_p} \exp(-\alpha_{k_p} k) \sin(\omega_{k_p} k + \theta_{k_p}) \\ \sum_{p=1}^{P} c_{k_p} \exp(-\alpha_{k_p} k) \cos(\omega_{k_p} k + \theta_{k_p}) \end{bmatrix} + \mathbf{v}_k \end{aligned} \tag{4.2}$$

for $k = 0, 1, \cdots, K-1$. The real and imaginary parts of the measurement are treated as separate measurements. $\mathbf{v}_k$ is a two element column vector containing the real and imaginary parts of the complex noise in (4.1) with $\mathbf{v}_k \sim N(0, R_k)$ where

$$R_k = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}. \tag{4.3}$$

The elements of the state variable vector $\mathbf{x}_k$ can be defined as

$$\begin{aligned} x_{k_{4(p-1)+1}} &= \omega_{k_p} \\ x_{k_{4(p-1)+2}} &= c_{k_p} \\ x_{k_{4(p-1)+3}} &= \theta_{k_p} \\ x_{k_{4(p-1)+4}} &= \alpha_{k_p}. \end{aligned} \tag{4.4}$$

The state variables obey the nonlinear plant equation

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{w}_k \tag{4.5}$$

where $\mathbf{w}_k \sim N(0, Q_k)$.

It is convenient and straightforward to treat the amplitudes (c_{k_p}), damping coefficients (α_{k_p}), frequencies (ω_{k_p}), and phases (θ_{k_p}) as state variables. This permits the time dependence of the signal parameters to be directly modeled through the process equation (4.5). An alternative approach is to model the differential equation for the measurements through the process equation. For the case of continuous time

real sinusoids let

$$y(t) = c(t)\exp(-\alpha(t)t)\sin(\omega(t)t + \theta(t))$$

then

$$\dot{y}(t) = -\alpha(t)\exp(-\alpha(t)t)\sin(\omega(t)t + \theta(t)) + \omega(t)\exp(-\alpha(t)t)\cos(\omega(t)t + \theta(t))$$

$$\ddot{y}(t) = (\alpha^2(t) - \omega^2(t))\exp(-\alpha(t)t)\sin(\omega(t)t + \theta(t))$$

$$- 2\alpha(t)\omega(t)\exp(-\alpha(t)t)\cos(\omega(t)t + \theta(t))$$

and the process equation can be expressed as

$$\dot{\mathbf{y}}(t) = \begin{bmatrix} \dot{y}(t) \\ \ddot{y}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -(\alpha^2(t) + \omega^2(t)) & -2\alpha(t) \end{bmatrix} \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix} + \mathbf{q}(t)$$

where $\mathbf{q}(t)$ is the process noise. The advantage of this formulation is that the measurement equation $z(t) = y(t) + v(t)$, where $v(t)$ is the process noise, is a linear function of the state. However, there are several disadvantages. The primary disadvantage is determining the initial conditions. Since the process equation may contain unknown parameters $c(t), \alpha(t), \omega(t)$, and $\theta(t)$, reasonably small error in the initial estimates of these parameters may lead to very poor initial estimates for $\dot{y}(0)$ and $\ddot{y}(0)$ causing the solution to converge to harmonics of the actual frequency or other poor filter performance. In addition, to express time dependence and initial uncertainty of the unknown parameters these parameters must also be modeled as state variables, thus further complicating the process equation. Even in the case where the unknown parameters are constant, the process equation is nonlinear. With these considerations it was decided that the models given by (4.5) and (4.2) were more appropriate and convenient for the harmonic retrieval problem.

4.1.1 Estimating the Parameters of 2 Sinusoids

A particular measurement model of interest consists of two exponentially damped sinusoids in white noise. This model has been analyzed by Kumaresan and Tufts [41] using reduced rank SVD techniques, and by Papadopoulos and Nikias [52] using cumulants.

For this model the measurement equation (4.2) becomes

$$\begin{aligned} \mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k \\ &= \begin{bmatrix} \sum_{p=1}^{2} \exp(-\alpha_{k_p} k) \sin(\omega_{k_p} k) \\ \sum_{p=1}^{2} \exp(-\alpha_{k_p} k) \cos(\omega_{k_p} k) \end{bmatrix} + \mathbf{v}_k. \end{aligned} \tag{4.6}$$

The associated state variables are defined as

$$\begin{aligned} x_{k_1} &= \omega_1 \\ x_{k_2} &= \alpha_1 \\ x_{k_3} &= \omega_2 \\ x_{k_4} &= \alpha_2 \end{aligned} \tag{4.7}$$

Assuming constant frequencies and damping coefficients, the plant equation that describes the evolution of the states (4.7) is given by

$$\mathbf{x}_{k+1_p} = \mathbf{x}_{k_p}. \tag{4.8}$$

4.2 Nonlinear Filters for Harmonic Retrieval

This section presents the equations used for the extended Kalman filter (EKF), the Gaussian second order(GSO) filter, the minimum variance filter (MVF), and three iterated forms of the extended Kalman filter for the harmonic retrieval

problem. The EKF and GSO equations are approximate solutions to the nonlinear estimation problem. The EKF requires the first order Taylor series expansion of the measurement about the latest estimate, and the GSO filter requires second order expansions. The MVF equations are exact expressions of the mear squared error state estimates for exponential nonlinearities in Gaussian noise. The three iterated filters are extensions of the EKF equations. The process model presented in (4.5) is appropriate for the general case of time varying parameters. However, in this chapter the state variables are restricted to be time invariant.

4.2.1 Extended Kalman Filter

The EKF is obtained by making Gaussian assumptions about the *a posteriori* densities and by extending the plant and measurement nonlinearities in a Taylor series including first order terms. The extended Kalman filter equations for time invariant states [6] (p. 195) are given by

$$\begin{aligned} K_k &= P_{k-1|k-1} H_k^T (H_k P_{k-1|k-1} H_k^T + R_k)^{-1} \\ P_{k|k} &= (I_n - K_k H_k) P_{k-1|k-1} \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k-1|k-1} + K_k \tilde{\mathbf{z}}_k \\ \tilde{\mathbf{z}}_k &= \mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) \end{aligned} \tag{4.9}$$

where $\hat{\mathbf{x}}_{k|k-1}$ is predicted estimate, $P_{k|k-1}$ is the one-step prediction covariance, K_k is the filter gain, $\hat{\mathbf{x}}_{k|k}$ is the filtered estimate, and $P_{k|k}$ is the filter covariance. I_n is the n-dimensional identity matrix. The filter requires the initial conditions $E[\mathbf{x}_0] = \hat{\mathbf{x}}_0$ and $E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)\,(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T] = P_0$.

For the measurement equation in (4.4) the Jacobian of $\mathbf{h}_k\mathbf{x}_k)$ is given by

$$
\begin{aligned}
H_k &= \left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}} \\
&= \left[\begin{array}{cc}
e^{-x_{k_2}k}\cos(x_{k_1}k) & -e^{-x_{k_2}k}\sin(x_{k_1}k) \\
-e^{-x_{k_2}k}\sin(x_{k_1}k) & -e^{-x_{k_2}k}\cos(x_{k_1}k) \\
e^{-x_{k_4}k}\cos(x_{k_3}k) & -e^{-x_{k_4}k}\sin(x_{k_3}k) \\
-e^{-x_{k_4}k}\sin(x_{k_3}k) & -e^{-x_{k_4}k}\cos(x_{k_3}k)
\end{array}\right]^T_{\mathbf{x}_k=\hat{\mathbf{x}}_{k-1|k-1}}
\end{aligned}
\tag{4.10}
$$

4.2.2 Gaussian Second Order Filter

The Gaussian second order (GSO) filter relations [6] are obtained by including second order terms in the Taylor series expansions of the plant and measurement equations. The constant state model of (4.4) leads to the GSO filter equations

$$
\begin{aligned}
K_k &= P_{k-1|k-1}H_k^T(H_kP_{k-1|k-1}H_k^T + R_k + B_k)^{-1} \\
\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + K_k\tilde{\mathbf{z}}_k \\
P_{k|k} &= (I_n - K_kH_k)P_{k-1|k-1}
\end{aligned}
\tag{4.11}
$$

where

$$
\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k-1|k-1}) - \frac{1}{2}\,\partial^2(\mathbf{h}_k, P_{k-1|k-1})\Big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}. \tag{4.12}
$$

and

$$
\partial^2(\mathbf{h}_k, P_{k-1|k-1}) = \text{trace}\left\{\left[\frac{\partial^2 g_{k_i}}{\partial x_p \partial x_q}\right]P_{k-1|k-1}\right\}. \tag{4.13}
$$

The bracketed quantity in (4.13) is a matrix whose pq^{th} element is the quantity $\frac{\partial^2 h_{k_i}}{\partial xp\,\partial xq}$. The matrix and B_k is approximated by

$$
B_{k_{ij}} \approx \frac{1}{4}\left[\sum_{p,q,m,n}\frac{\partial^2 h_{k_j}}{\partial x_p\,\partial x_q}(d_{pm}d_{qn} + d_{pn}d_{qm})\frac{\partial^2 h_{k_i}}{\partial x_m\,\partial x_n}\right]_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}} \tag{4.14}
$$

where h_{k_i} denotes the i^{th} element of $\mathbf{h}_k(.)$, and the d's are elements of $P_{k-1|k-1}$.

For the model (4.8)

$$\begin{aligned}\frac{\partial^2 h_{k_1}(\mathbf{x}_k)}{\partial \mathbf{x}_k \partial \mathbf{x}_k^T} &= k^2 \begin{bmatrix} e^{-x_{k_2}k} D_{k_1} & 0 \\ 0 & e^{-x_{k_4}k} D_{k_3} \end{bmatrix} \\ \frac{\partial^2 h_{k_2}(\mathbf{x}_k)}{\partial \mathbf{x}_k \partial \mathbf{x}_k^T} &= k^2 \begin{bmatrix} e^{-x_{k_2}k} E_{k_1} & 0 \\ 0 & e^{-x_{k_4}k} E_{k_3} \end{bmatrix}\end{aligned} \tag{4.15}$$

where

$$\begin{aligned} D_{k_i} &= \begin{bmatrix} -\sin(x_{k_i}k) & -\cos(x_{k_i}k) \\ -\cos(x_{k_i}k) & \sin(x_{k_i}k) \end{bmatrix} \\ E_{k_i} &= \begin{bmatrix} -\cos(x_{k_i}k) & \sin(x_{k_i}k) \\ \sin(x_{k_i}k) & \cos(x_{k_i}k) \end{bmatrix} \end{aligned} \tag{4.16}$$

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1}) + \frac{1}{2}\,\partial^2(\mathbf{f}_{k-1}, P_{k-1|k-1})\Big|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}} \\ P_{k|k-1} &= F_{k-1} P_{k-1|k-1} F_{k-1}^T + \Gamma_{k-1} Q_{k-1} \Gamma_{k-1}^T + A_{k-1} \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + K_k \tilde{\mathbf{z}}_k \\ P_{k|k} &= (I_n - K_k H_k) P_{k|k-1} \\ K_k &= P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k + B_k)^{-1}. \end{aligned} \tag{4.17}$$

4.2.3 Minimum Variance Filter

The EKF and the GSO filters are based on a Taylor series expansion of the nonlinear equations about the most recent estimate. As such the EKF and GSO filters are subject to the inherent problems of local linearizations and may lead to poor performance. Liang [23] developed a minimum variance filter (MVF) which gives exact estimates at each iteration of the filter based on the assumption

that the estimation errors are Gaussian. He has shown that for certain nonlinear functions such as polynomial nonlinearities, exponential functions, and sinusoids, exact expressions for the state estimates can be obtained and used in the filter relations in place of the usual approximations. At each step in the operation they assume that the prediction and filter errors are Gaussian. They have compared their filter to the EKF and other filters using numerical examples and claim that their filter performs much better than the EKF for large initial error variances. Using the plant and measurement models in (4.5) and (4.6) the minimum variance filter relations are given by

$$\begin{aligned} K_k &= E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{h}}_k(\mathbf{x}_k)^T]\,(R_k + E[\tilde{\mathbf{h}}_k(\mathbf{x}_k)\tilde{\mathbf{h}}_k(\mathbf{x}_k)^T])^{-1} \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k-1|k-1} + K_k\tilde{\mathbf{z}}_k = \mathbf{z}_k - \hat{\mathbf{h}}_k(\mathbf{x}_k) \\ P_{k|k} &= P_{k-1|k-1} - K_k\,E[\tilde{\mathbf{h}}_k(\mathbf{x}_k)\tilde{\mathbf{x}}_{k|k-1}^T] \end{aligned} \tag{4.18}$$

where

$$\begin{aligned} \tilde{\mathbf{x}}_{k|k-1} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \\ \tilde{\mathbf{h}}_k(\mathbf{x}_k) &= \mathbf{h}_k(\mathbf{x}_k) - \hat{\mathbf{h}}_k(\mathbf{x}_k) \\ \tilde{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1}) &= \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) - \hat{\mathbf{f}}_{k-1}(\mathbf{x}_{k-1}). \end{aligned} \tag{4.19}$$

The MVF requires exact analytical expressions for $E[\mathbf{h}_k(\mathbf{x}_k)]$, $E[\mathbf{x}_k\mathbf{h}_k(\mathbf{x}_k)^T]$, and $E[\mathbf{h}_k(\mathbf{x}_k)\mathbf{h}_k(\mathbf{x}_k)^T]$.

The general system model (4.2) requires closed form relations for functions of the form $E[\exp(y_k)]$, $E[x_{k_i}\exp(y_k)]$, and $E[x_{k_i}\,x_{k_j}\exp(y_k)]$, where y_k is defined by the inner product

$$y_k = \mathbf{u}_k^T\mathbf{x}_k$$

and where $\mathbf{u}_k$ is a vector of deterministic coefficients. Note that if $\mathbf{x}_k$ is a vector of jointly Gaussian variables, then y_k is also Gaussian. Liang [24] derives relations

for expectations of this form. The following relations can be used to evaluate these expectations for the general system model given in (4.2) :

$$
\begin{aligned}
E[\exp(y_k)] &= \exp(\mathbf{u}_k^T \hat{\mathbf{x}}_{k|k-1} + \frac{1}{2}\mathbf{u}_k^T P_{k|k-1}\mathbf{u}_k) \\
E[x_{k_i}\exp(y_k)] &= (\hat{x}_{k|k-1_i} + \mathbf{u}_k^T P_{k|k-1}\mathbf{e}_i) \\
&\quad \times \exp(\mathbf{u}_k^T \hat{\mathbf{x}}_{k|k-1} + \frac{1}{2}\mathbf{u}_k^T P_{k|k-1}\mathbf{u}_k) \\
E[x_{k_i} x_{k_j} \exp(y_k)] &= (\mathbf{e}_i^T P_{k|k-1}\mathbf{e}_j + \hat{x}_{k|k-1_i}\hat{x}_{k|k-1_j} \\
&\quad + \hat{x}_{k|k-1_i}\mathbf{u}_k^T P_{k|k-1}\mathbf{e}_j + \hat{x}_{k|k-1_j}\mathbf{u}_k^T P_{k|k-1}\mathbf{e}_i \\
&\quad + \mathbf{u}_k^T P_{k|k-1}\mathbf{e}_j \mathbf{u}_k^T P_{k|k-1}\mathbf{e}_i) \\
&\quad \times \exp(\mathbf{u}_k^T \hat{\mathbf{x}}_{k|k-1} + \frac{1}{2}\mathbf{u}_k^T P_{k|k-1}\mathbf{u}_k)
\end{aligned}
\tag{4.20}
$$

where $\mathbf{e}_i$ is the i^{th} unit vector. This vector is zero except for the i^{th} element.

For the measurement equation in (4.6), in which the amplitudes of the sinusoids c_{k_p} are known, only the first two expressions in (4.20) are necessary.

4.2.3.1 Evaluation of $E[\mathbf{h}_k]$

Let the measurement nonlinearity from (4.6) be expressed as

$$
\mathbf{h}_k = \begin{bmatrix} \mathrm{Im}\left(\exp(\mathbf{u}_1^T \mathbf{x}_k) + \exp(\mathbf{u}_2^T \mathbf{x}_k)\right) \\ \mathrm{Re}\left(\exp(\mathbf{u}_1^T \mathbf{x}_k) + \exp(\mathbf{u}_2^T \mathbf{x}_k)\right) \end{bmatrix} \tag{4.21}
$$

where

$$
\begin{aligned}
\mathbf{u}_{k_1}^T &= [\ jk \quad -k \quad 0 \quad 0\ ] \\
\mathbf{u}_{k_2}^T &= [\ 0 \quad 0 \quad jk \quad -k\ ]
\end{aligned}
\tag{4.22}
$$

Define the quantities

$$\begin{aligned}\hat{h}_{k_1} &= \exp(\mathbf{u}_{k_1}^T \hat{\mathbf{x}}_{k|k-1} + \frac{1}{2}\mathbf{u}_{k_1}^T P_{k|k-1}\mathbf{u}_{k_1}^T) \\ \hat{h}_{k_2} &= \exp(\mathbf{u}_{k_2}^T \hat{\mathbf{x}}_{k|k-1} + \frac{1}{2}\mathbf{u}_{k_2}^T P_{k|k-1}\mathbf{u}_{k_2}^T)\end{aligned} \tag{4.23}$$

Using (4.23) and (4.20) the expression for the expected value of the measurements for the model in (4.6) is obtained as

$$E[\mathbf{h}_k] = \begin{bmatrix} \mathrm{Im}(\hat{h}_{k_1} + \hat{h}_{k_2}) \\ \mathrm{Re}(\hat{h}_{k_1} + \hat{h}_{k_2}) \end{bmatrix}. \tag{4.24}$$

4.2.3.2 Evaluation of $E[\mathbf{x}_k\, \mathbf{h}_k]^T$

For the model in (4.6), $\mathrm{E}[\mathbf{x}_k \mathbf{h}_k^T]$ can be found from (4.20) using

$$\begin{aligned}\hat{\mathbf{C}}_k &= \mathrm{E}[\mathbf{x}_k\, (\exp(\mathbf{u}_{k_1}^T \mathbf{x}_k) + \exp(\mathbf{u}_{k_2}^T \mathbf{x}_k)] \\ &= (\hat{\mathbf{x}}_{k|k-1} + \mathbf{u}_{k_1}^T P_{k|k-1})\hat{h}_{k_1} + (\hat{\mathbf{x}}_{k|k-1} + \mathbf{u}_{k_2}^T P_{k|k-1})\hat{h}_{k_2}\end{aligned} \tag{4.25}$$

which gives

$$\mathrm{E}[\mathbf{x}_k \mathbf{h}_k^T] = [\mathrm{Im}(\hat{\mathbf{C}}_k) \vdots \mathrm{Re}(\hat{\mathbf{C}}_k)] \tag{4.26}$$

4.2.3.3 Evaluation of $E[\mathbf{h}_k\, \mathbf{h}_k]^T$

$\mathrm{E}[\mathbf{h}_k\, \mathbf{h}_k^T]$ for the model in (4.6) is evaluated using the first equation in (4.20).

$$E[\mathbf{h}_k\, \mathbf{h}_k^T] = \begin{bmatrix} \hat{h}_{k_{11}} & \hat{h}_{k_{12}} \\ \hat{h}_{k_{12}} & \hat{h}_{k_{22}} \end{bmatrix} \tag{4.27}$$

where

$$\begin{aligned}\hat{h}_{k_{11}} &= \mathrm{Re}\{[1 \quad -1 \quad 2 \quad -2 \quad 1 \quad 1]\,\mathbf{E}_k/2\} \\ \hat{h}_{k_{12}} &= \mathrm{Im}\{[0 \quad 1 \quad 0 \quad 2 \quad 0 \quad 1]\,\mathbf{E}_k/2\} \\ \hat{h}_{k_{22}} &= \mathrm{Re}\{[1 \quad 1 \quad 2 \quad 2 \quad 1 \quad 1]\,\mathbf{E}_k/2\}\end{aligned} \tag{4.28}$$

The i^{th} element of the column vector $\mathbf{E_k}$ is defined as

$$\begin{aligned} E_{k_i} &\equiv \mathrm{E}[\exp(\mathbf{U}_{k_i}\hat{\mathbf{x}}_k] \\ &\equiv \exp(\mathbf{U}_{k_i}\hat{\mathbf{x}}_{k|k-1} + \mathbf{U}_{k_i}P_{k|k-1}\mathbf{U}_{k_i}^T/2) \end{aligned} \tag{4.29}$$

and where $\mathbf{U}_{k_i}$ is the i^{th} of the matrix U_k, which is given by

$$U_k = k \begin{bmatrix} 0 & -2 & 0 & 0 \\ 2j & -2 & 0 & 0 \\ j & -1 & -j & -1 \\ j & -1 & j & -1 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 2j & -2 \end{bmatrix} \tag{4.30}$$

The quantities $E[\mathbf{h}_k]$, $E[\mathbf{x}_k\,\mathbf{h}_k^T]$, and $E[\mathbf{h}_k\,\mathbf{h}_k^T]$, found from (4.24), (4.26), and (4.27), respectively, are used in (4.18) to form the minimum variance filter expressions.

4.2.4 Iterated Filters

Three iterated forms of the extended Kalman filter are also used for parameter estimation. The iterated filters can be categorized into two classes. Locally iterated filters are implemented by continuously processing the data for a given measurement (i.e. for a given value of k) until the error between iterations is minimized or until a maximum iteration count is exceeded. Globally iterated filters are implemented by processing the entire data set more than once essentially recycling the data set through the filter.

4.2.4.1 Locally Iterated Extended Kalman Filter

The locally iterated Kalman filter (LIKF) is an enhanced version of the extended Kalman filter where, at each step of the iteration procedure, the measurement nonlinearity is linearized about the state estimate obtained from the EKF equations. This filter was first introduced by Denham and Pines [21]. The procedure is to repetitively calculate $\hat{\mathbf{x}}_{k|k}, K_k$, and $P_{k|k}$, each time by linearizing about the most recent estimate. The recursion relations for the LIKF are given by [6] (p. 190)

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k}(i+1) &= \hat{\mathbf{x}}_{k|k-1} + K_k(i)[\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k}(i)) - H_k(\hat{\mathbf{x}}_{k|k-1} - \hat{\mathbf{x}}_{k|k}(i))] \\
P_{k|k}(i) &= (I_n - K_k(i)H_k)P_{k|k-1} \\
K_k(i) &= P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + R_k)^{-1}
\end{aligned} \tag{4.31}$$

where $i = 0, 1, \cdots$. The number of repetitions of the calculations shown above can be determined by requiring the magnitude of the difference between successive state estimates to be less than some small number.

4.2.4.2 Globally Iterated Extended Kalman Filter

Another form of the iterated Kalman filter, designated the globally iterated extended Kalman filter (GIKF) [6], involves restarting the filter after each complete pass through the data. After filtering the K measurements with the extended Kalman filter the covariance is reset back to its initial value and the filter is restarted with the first measurement but using the final estimate from the previous iteration as the new initial estimate. This technique can be repeated until the difference in the final estimates from successive iterations converges to some small value. By resetting the covariance the system is essentially re-excited thereby allowing the state estimates to be perturbed. The premise for this technique is that good estimates

will not be made worse, but poor estimates may be forced to better values.

4.2.4.3 Covariance Resetting

A similar procedure can be applied within a single pass of the data. For example whenever the state covariance converges rapidly to a relatively small steady state value, resetting of the covariance to a point between its present value and the initial value takes place. The effect of resetting is to re-excite the system after steady state is reached, so that early saturation of the filter gain to a small value that would prevent changes in the estimates is avoided. The disadvantage of this technique is that the variance of the final estimates could not be as good as it would if no resetting took place. The advantage is that poor estimates may be forced to better values. This technique could also be done through multiple passes of the same data. During the final pass the covariance would not be reset within the single pass, thus allowing the estimates to converge to the best possible values. This filtering technique will be referred to as the extended Kalman filter with resetting (EKFR).

4.2.5 Experimental Results for Estimation in White Noise

A well documented problem that has been traditionally approached using AR-based techniques is the estimation of the damping coefficients and frequencies of two sinusoids [41, 52]. Referring to equation (4.4) damping coefficient values were $\alpha_{k_1} = 0.2$ and $\alpha_{k_2} = 0.1$ with normalized frequencies of $\omega_{k_1} = 0.42 * 2\pi$ radians and $\omega_{k_2} = 0.52 * 2\pi$ radians for $0 \leq k \leq 24$.

The six nonlinear filters discussed previously are applied to this problem. These filters will be designated as: EKF - extended Kalman filter, GSO - Gaussian second order filter, MVF - minimum variance filter, LIKF - locally iterated extended

Kalman filter, GIKF - globally iterated extended Kalman filter, and EKFR - extended Kalman filter with covariance resetting. In the implementation of the LIKF, using (4.31) the repetition cycle for a given sample was terminated whenever either the magnitude of the difference between successive state estimates was less that 0.0001 or whenever a total of 9 repetitions were completed. The GIKF was implemented by processing the entire set of measurements five times. The EKFR was implemented by resetting the covariance every three measurements. The covariance was reset according to the formula

$$P_{k|k}(-) = \frac{P_{\text{ref}} + P_{k-1|k-1}}{2} \quad k = 2, 5, 8, \cdots \tag{4.32}$$

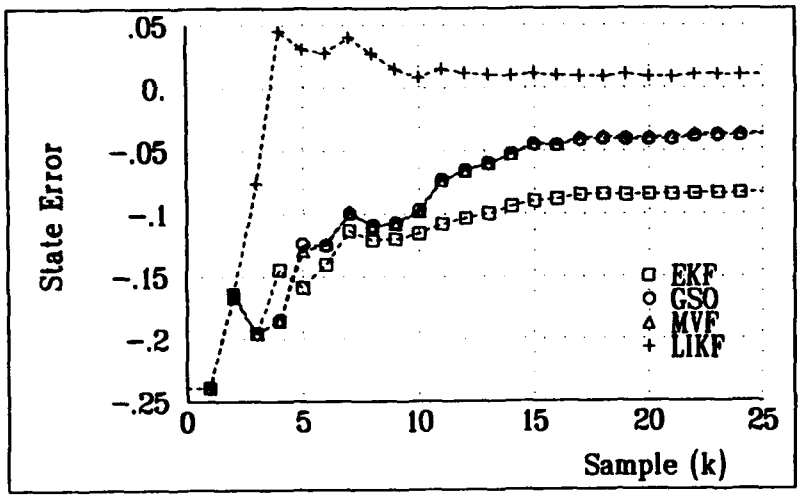
where initially $P_{\text{ref}} = P_0$. After $P_{k|k}(-)$ is computed the new reference becomes $P_{\text{ref}} = P_{k|k}(-)$. Using the EKFR the covariance was reset during the first two passes using (4.32). On the third and final pass the covariance was not reset. At the beginning of each of the second and third passes initial condition on the state was set to the final state estimate from the previous pass.

The filter performance was evaluated as a function of signal-to-noise ratio (SNR) for a range of 0 dB to 30 dB. SNR is defined as $10\log\frac{1}{2\sigma^2}$, where σ^2 is the variance of each of the real and imaginary components of the i.i.d. complex noise. Note that this definition gives the peak SNR. Performance was evaluated at each SNR by forming the sample variance of the estimation error over 500 independent noise runs. In each run the signal was kept the same while the noise was modified using different random number seeds.

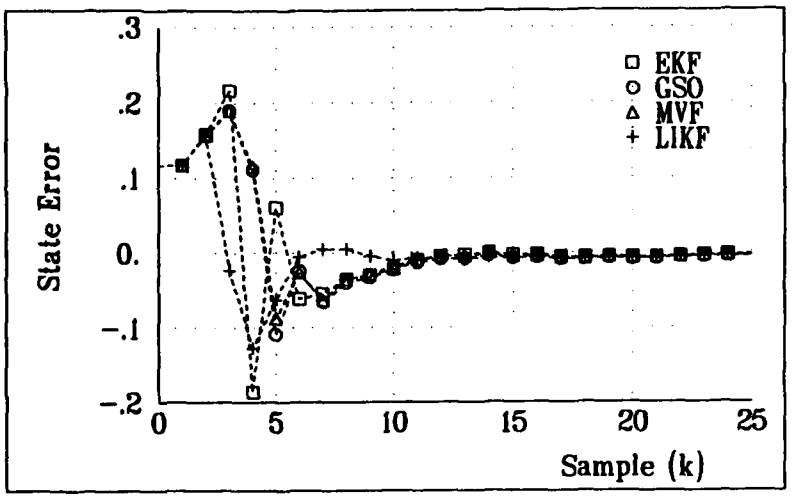
Figure 4.1 illustrates the estimation error as a function of sample number for a representative run at 20 dB SNR. This figure compares the relative performance of the non-iterated filters (the EKF, GSO, MVF) and the locally iterated LIKF. The

diagonal elements of the initial covariance were set to a value of 0.04 and the initial errors were randomly chosen based on this value. Overall this figure shows that the LIKF outperforms any of the noniterated filters. The MVF and the GSO give about the same results, and the EKF gives the worst performance. In this example the filters perform about the same in estimating the other state variables. Figure 4.2 shows the diagonal elements of the filter covariance $P_{k|k}$ for the same example. This graph shows that the covariance elements for the LIKF converge more rapidly than those for the other filters. It can be seen from this example that at 20 dB SNR the filter converges to its final state estimate within about 10 samples. Figures 4.3 and 4.4 show the estimation error and sample covariances, respectively, at 0 dB SNR. These figures again demonstrate that LIKF generally outperforms the other estimators, and that the EKF performs the worst. At 0 dB SNR the estimates stabilize in about $12 - 15$ samples for $P_0 = 0.04$. This illustrates the suitability of using these nonlinear filtering techniques for short data lengths.

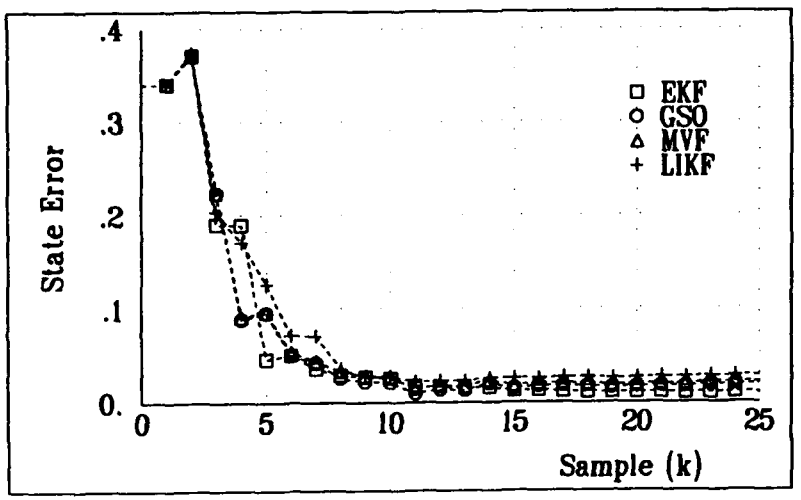
Figure 4.5 presents the performance results of the three noniterative filters as a function of SNR for $P_0 = 0.01$. Each point on graph represents the sample variance of the estimation error over 500 simulation runs. The results of the Kumaresan-Tufts (KT) method and fourth order cumulant (FOC) method obtained from [52] are also shown. The Cramer-Rao bound is also shown. This bound is derived in the Appendix. Figure 4.5 illustrates that all of the noniterative filters give similar performance for $P_0 = 0.01$. The performance is very close to the CR bound, particularly at high SNR's. The results for the first sinusoid ($\omega_{k_1} = 0.42 * 2\pi, \alpha_{k_1} = 0.2$) are slightly worse than those for the second sinusoid ($\omega_{k_2} = 0.52 * 2\pi, \alpha_{k_2} = 0.1$). This is probably because the first signal has been damped significantly before the filter has converged. The nonlinear filter results are significantly better than the KT



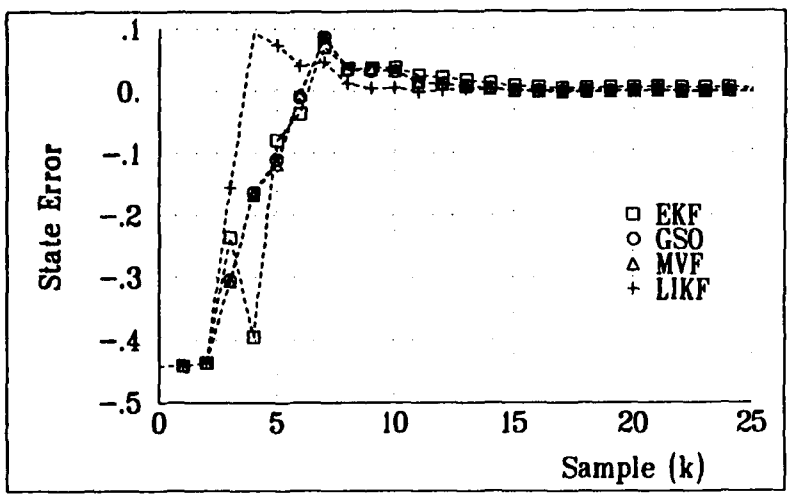
(a) Estimation Error for $f_{k_1} = 0.42$



(c) Estimation Error for $f_{k_2} = 0.52$

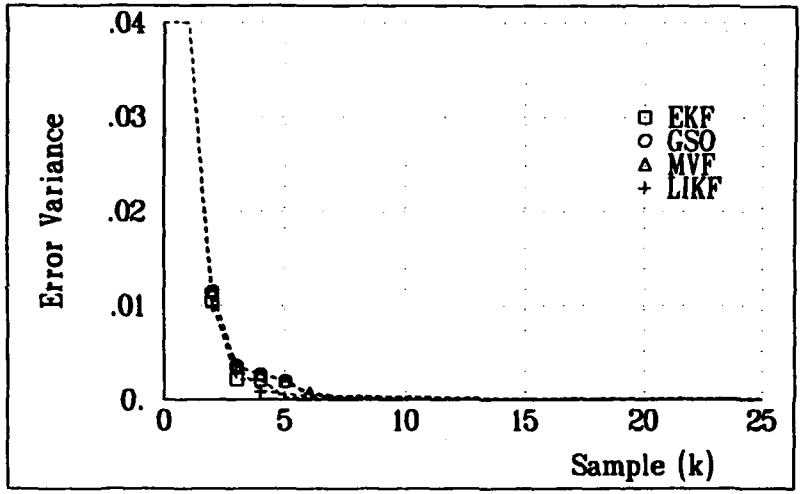


(b) Estimation Error for $\alpha_{k_1} = 0.2$

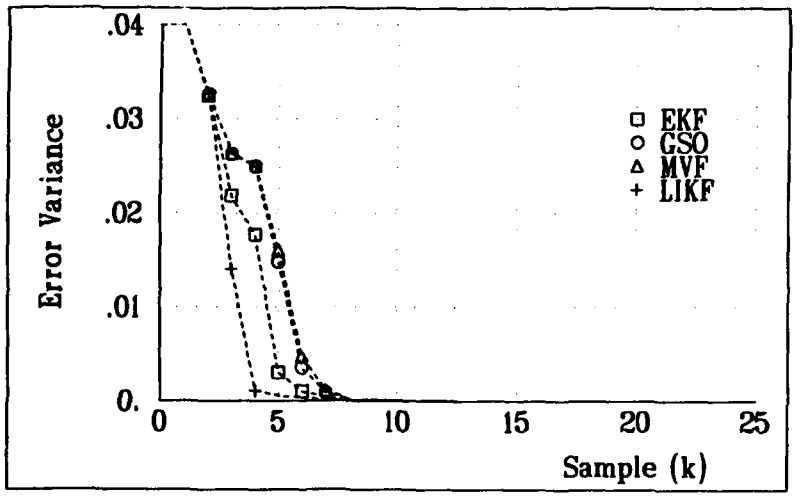


(d) Estimation Error for $\alpha_{k_2} = 0.1$

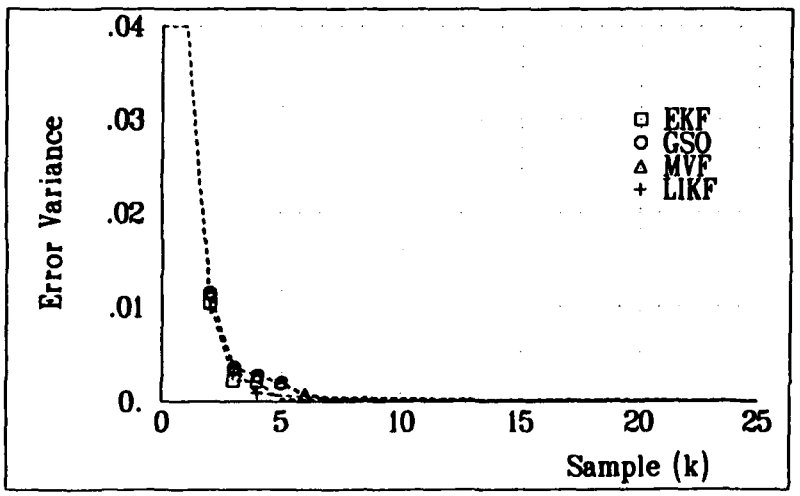
Figure 4.1 Typical Simulation Results for the Estimation Error, 4-State Model, SNR= 20 dB, $P_0 = 0.04\ I$



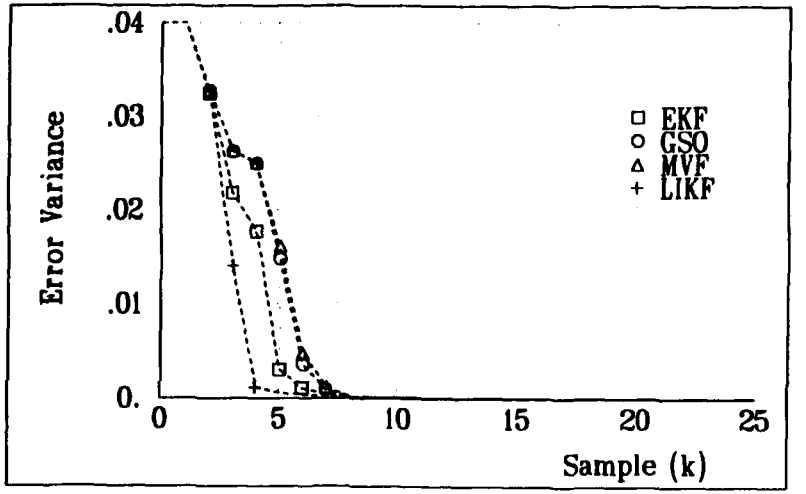
(a) Variance $P_{k|k}$ for $f_{k_1} = 0.42$



(c) Variance $P_{k|k}$ for $f_{k_2} = 0.52$

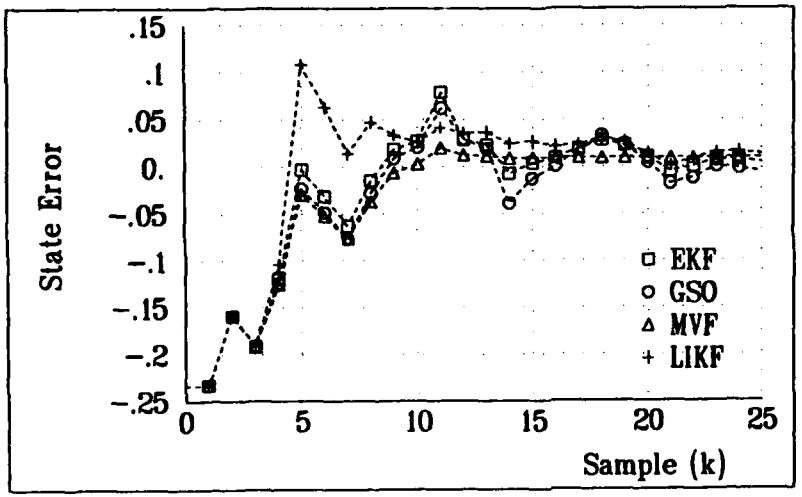


(b) Variance $P_{k|k}$ for $\alpha_{k_1} = 0.2$

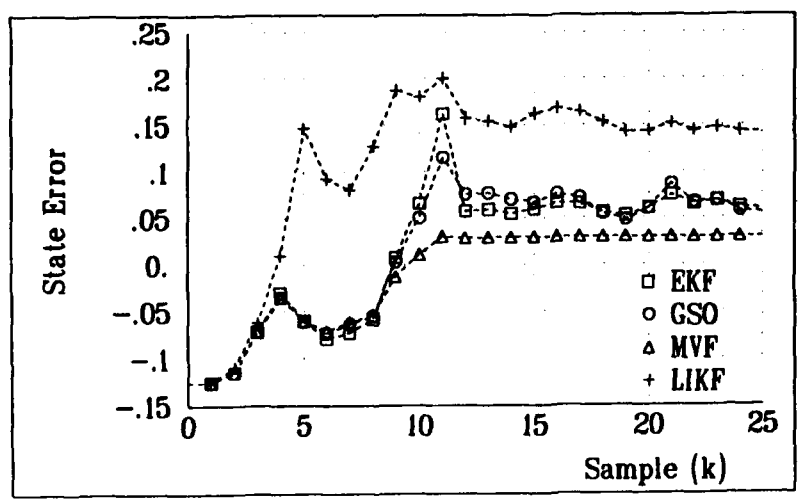


(d) Variance $P_{k|k}$ for $\alpha_{k_2} = 0.1$

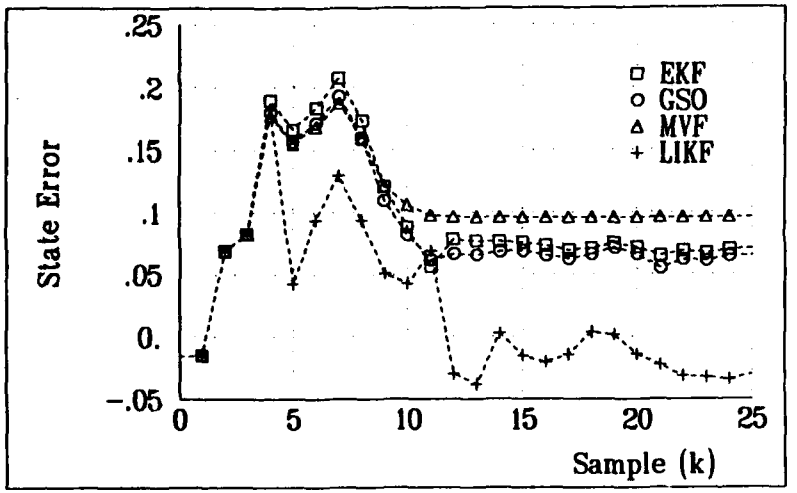
Figure 4.2 Typical Simulation Results for the Error Variance, 4-State Model, SNR= 20 dB, $P_0 = 0.04\ I$



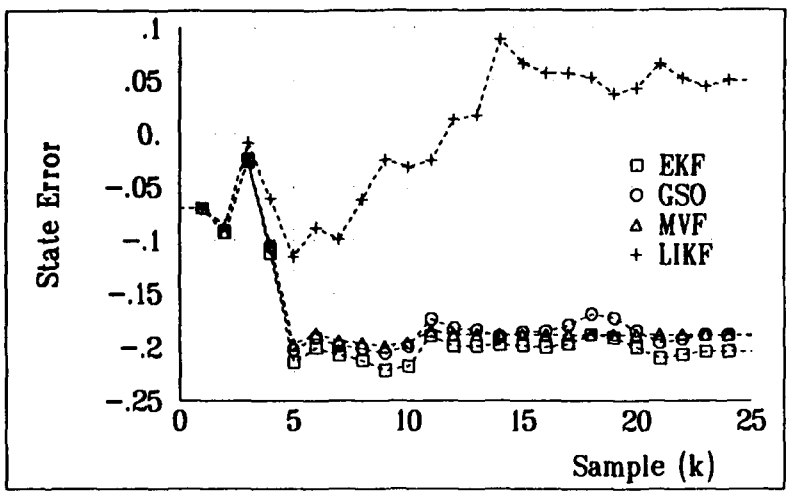
(a) Estimation Error for $f_{k_1} = 0.42$



(b) Estimation Error for $\alpha_{k_1} = 0.2$

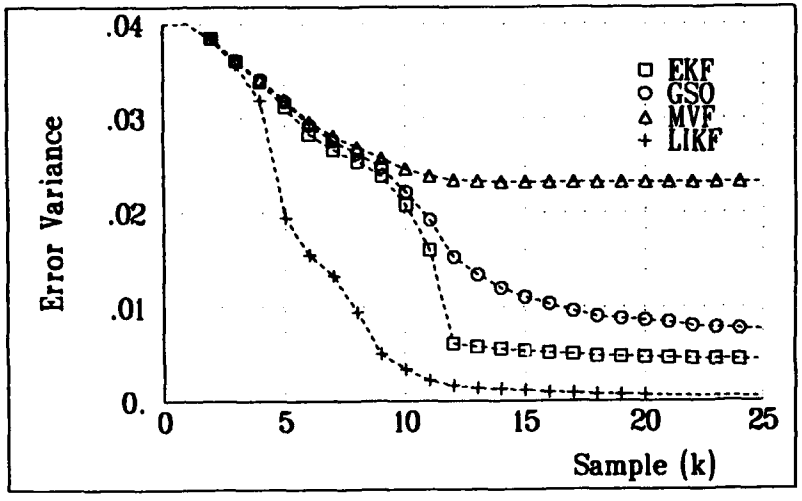


(c) Estimation Error for $f_{k_2} = 0.52$

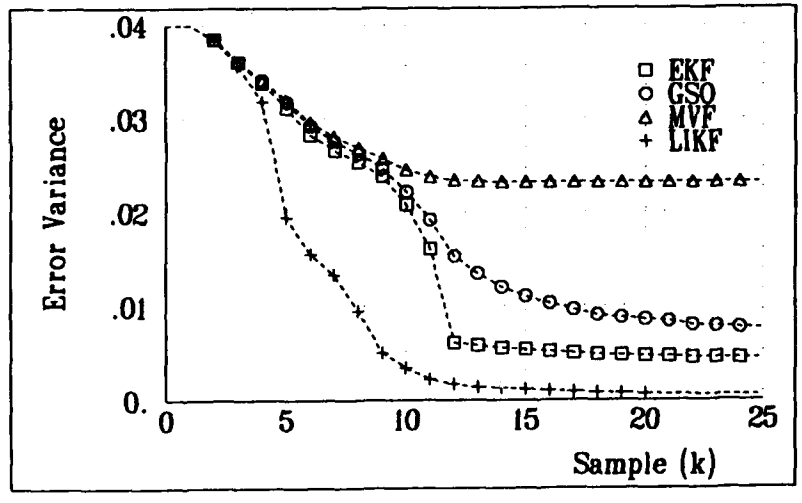


(d) Estimation Error for $\alpha_{k_2} = 0.1$

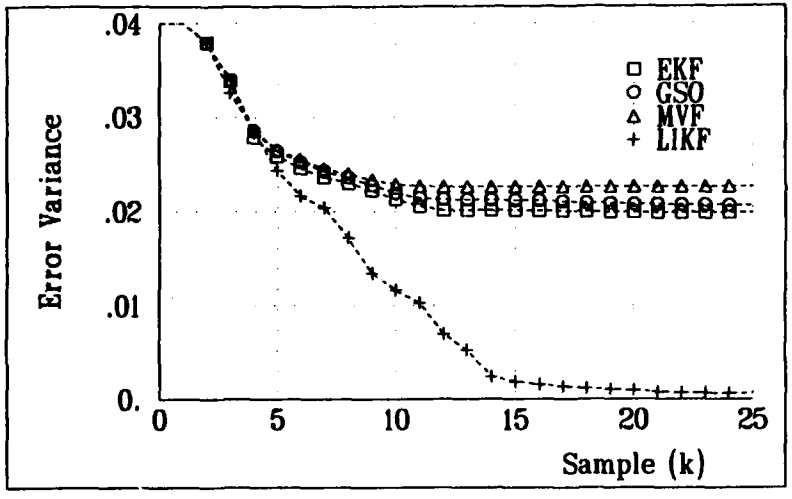
Figure 4.3 Typical Simulation Results for the Estimation Error, 4-State Model, SNR= 0 dB, $P_0 = 0.04\ I$



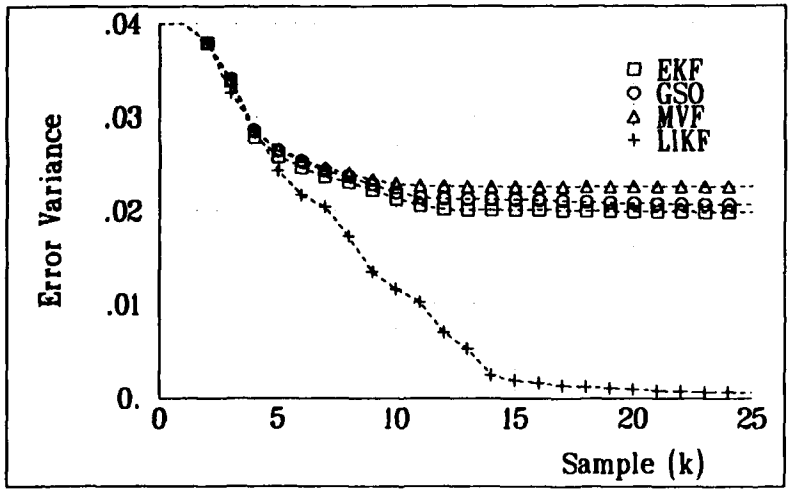
(a) Variance $P_{k|k}$ for $f_{k_1} = 0.42$



(b) Variance $P_{k|k}$ for $\alpha_{k_1} = 0.2$



(c) Variance $P_{k|k}$ for $f_{k_2} = 0.52$



(d) Variance $P_{k|k}$ for $\alpha_{k_2} = 0.1$

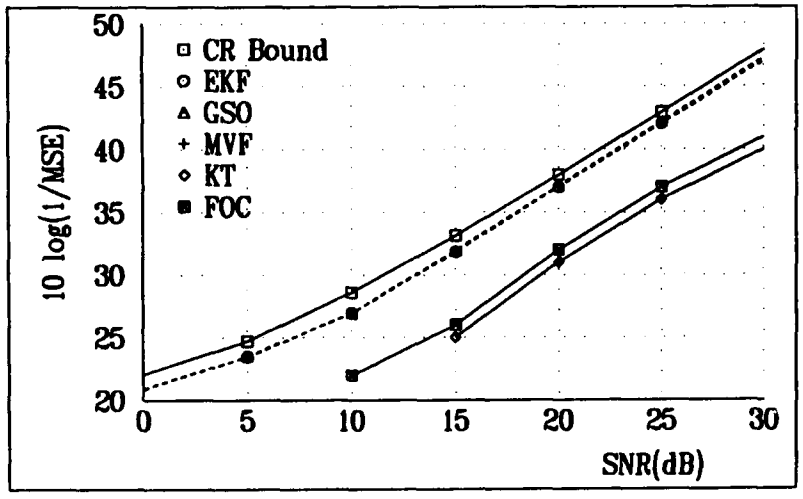
Figure 4.4 Typical Simulation Results for the Error Variance, 4-State Model, SNR= 0 dB, $P_0 = 0.04\ I$

and FOC results. However the KT method makes no assumptions about the noise statistics and does not require initial estimates. The nonlinear filtering techniques assume that the noise statistics are known *a priori*. Section 4.4 presents a technique to estimate the noise statistics on line.

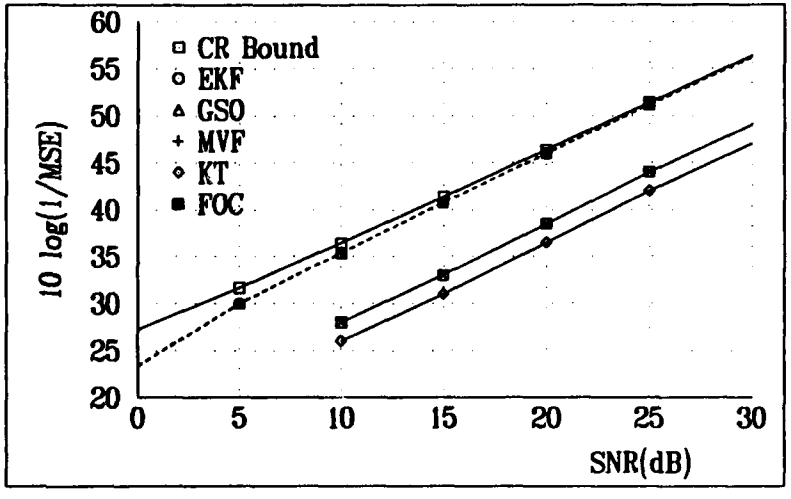
The results of the iterative techniques (LIKF, GIKF, EKFR) are given in Figure 4.6. These results are slightly better than those for the noniterative filters for SNR's above 10 dB, but slightly inferior for SNR's below 10 dB. Of course the price paid for better performance is higher computational requirements. Among the three noniterative filters the LIKF performs slightly better than the other two, particularly at low SNR's.

As the initial covariance increases the filter performance degrades. This is illustrated in Figures 4.7 and 4.8, where the mean squared estimation error as a function of SNR is shown for the iterative and noniterative filters, respectively, for $P_0 = 0.04$. The higher order forms of the noniterative filters, the GSO and the MVF, perform better than the EKF especially at high SNR's. This satisfies intuition in that the second order approximations for the GSO and the assumptions of Gaussian error distribution for the MVF should be more valid at high SNR's than at low SNR's. Results for all filters are still better than the KT and FOC methods. However again it must be emphasized that the KT method makes no assumptions about the noise statistics or the initial estimation error.

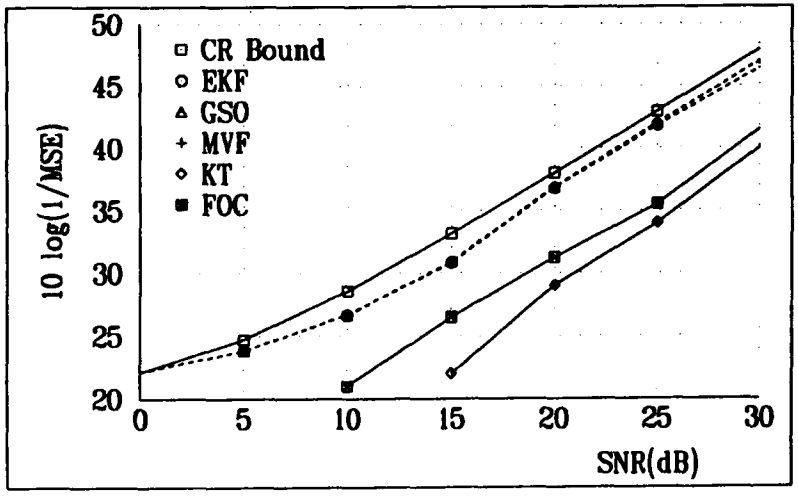
Least squares-based techniques such as the KT method generally fall apart at around 15 dB SNR due to the so-called threshold effect. A threshold occurs whenever there are more than P zeros outside the unit circle in the singular value decomposition process. One of the advantages of using filtering methods is that



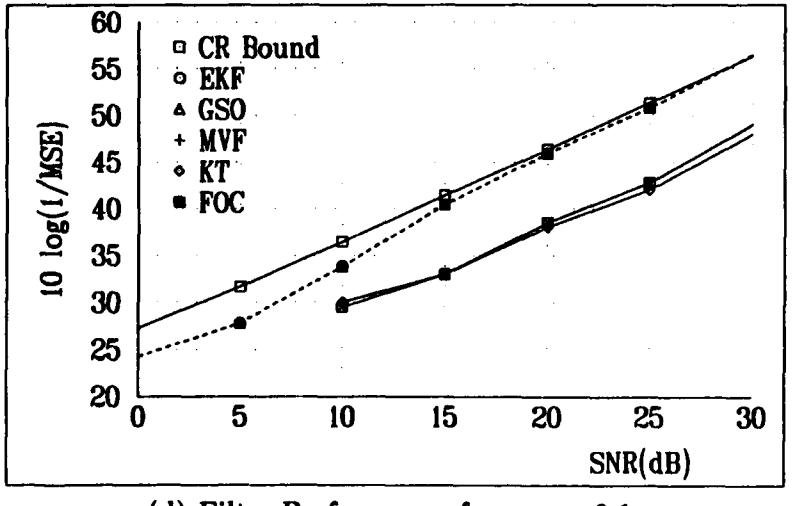
(a) Filter Performance for $\omega_{k_1} = 0.42 * 2\pi$



(c) Filter Performance for $\omega_{k_2} = 0.52 * 2\pi$

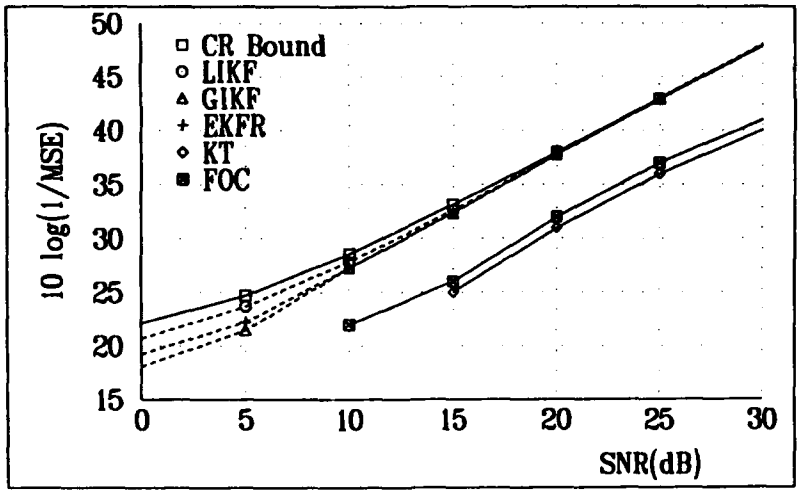


(b) Filter Performance for $\alpha_{k_1} = 0.2$

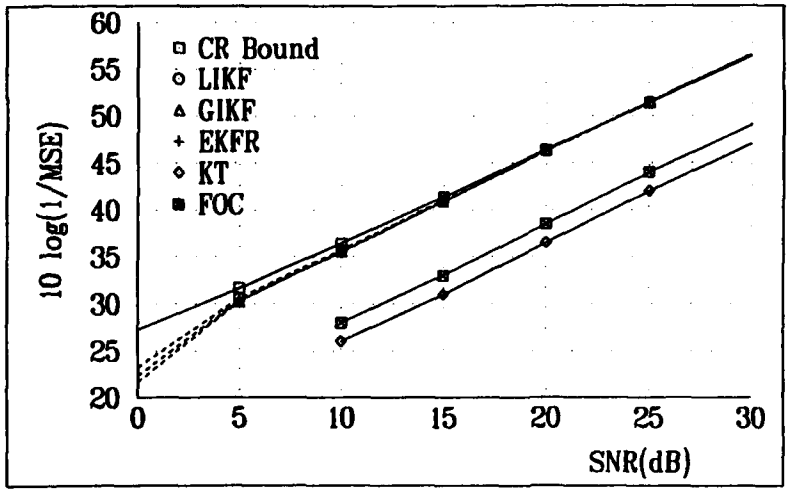


(d) Filter Performance for $\alpha_{k_2} = 0.1$

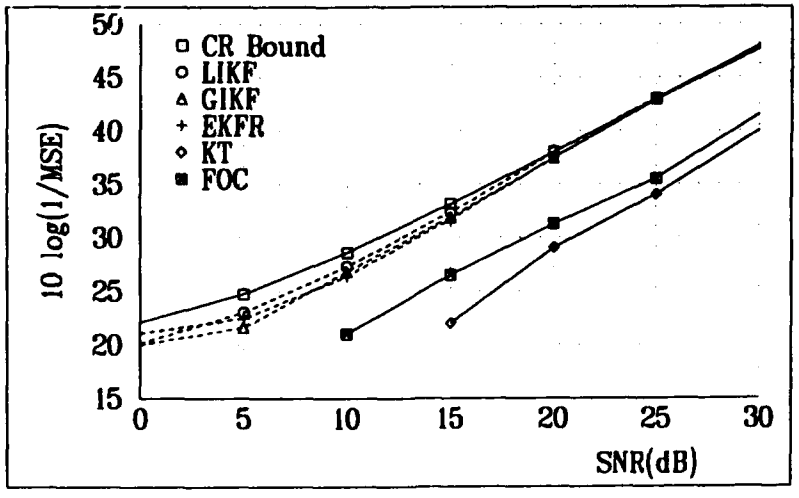
Figure 4.5 Noniterated Filter Performance in White Noise, 4-State Model, $P_0 = 0.01\ I$



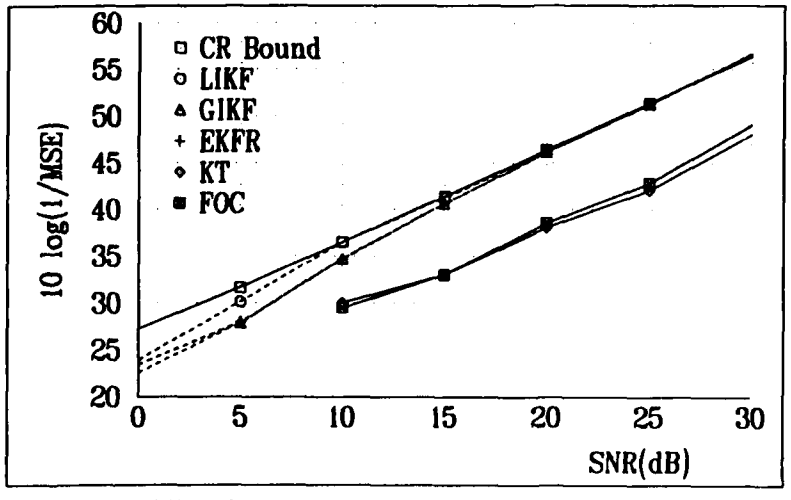
(a) Filter Performance for $\omega_{k_1} = 0.42 * 2\pi$



(c) Filter Performance for $\omega_{k_2} = 0.52 * 2\pi$

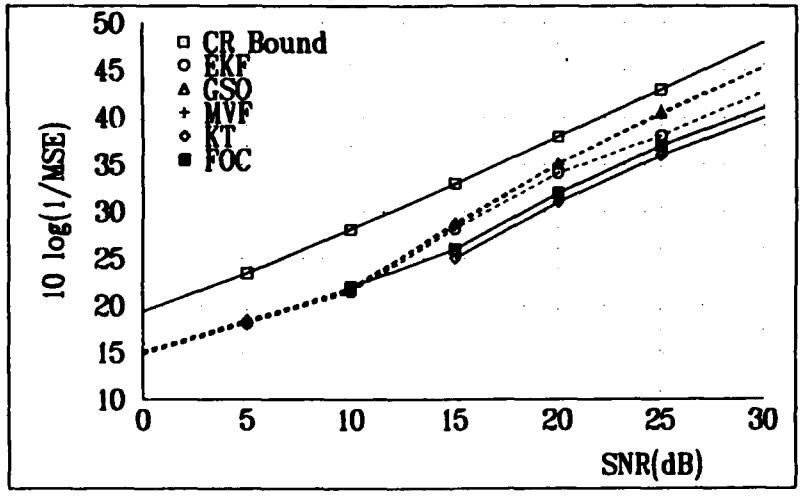


(b) Filter Performance for $\alpha_{k_1} = 0.2$

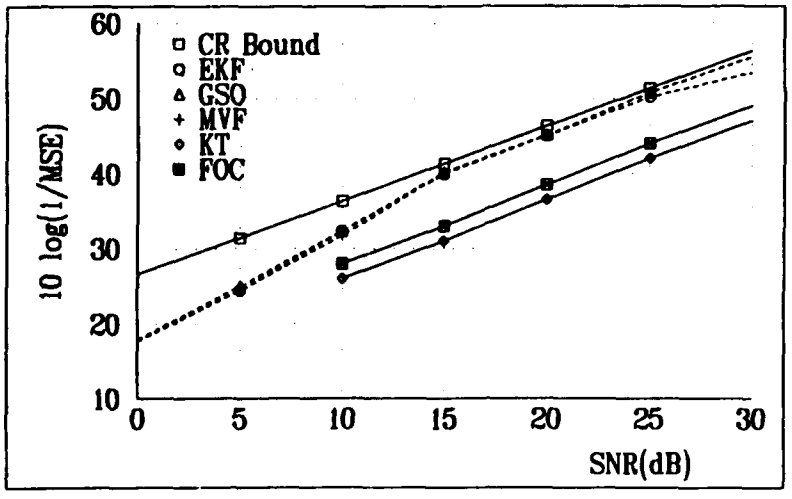


(d) Filter Performance for $\alpha_{k_2} = 0.1$

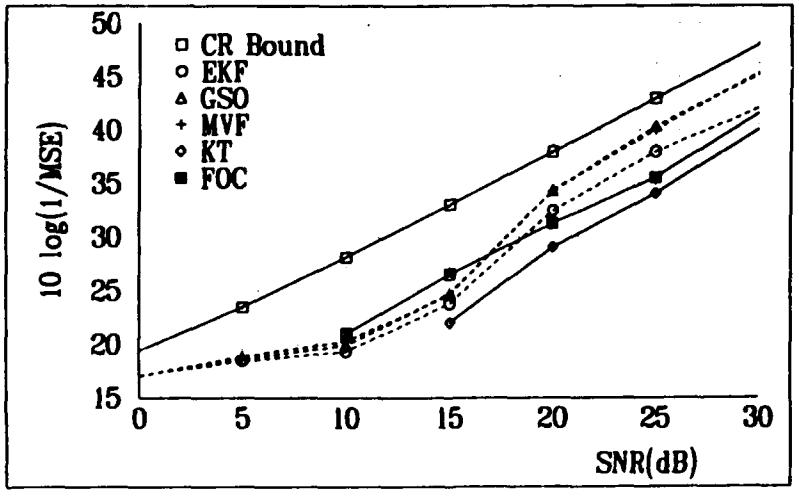
Figure 4.6 Iterated Filter Performance in White Noise, 4-State Model, $P_0 = 0.01\ I$



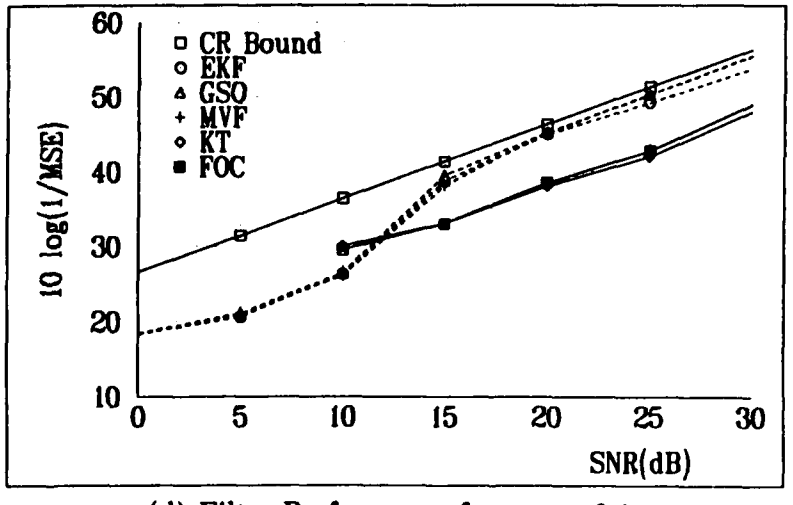
(a) Filter Performance for $\omega_{k_1} = 0.42 * 2\pi$



(c) Filter Performance for $\omega_{k_2} = 0.52 * 2\pi$

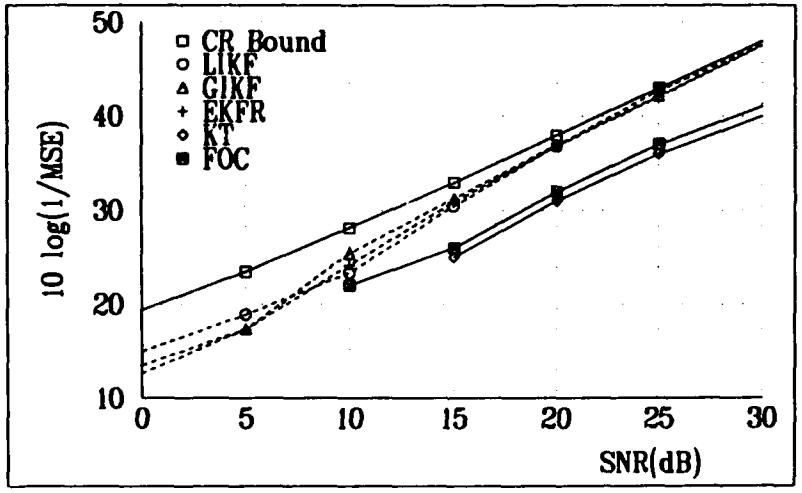


(b) Filter Performance for $\alpha_{k_1} = 0.2$

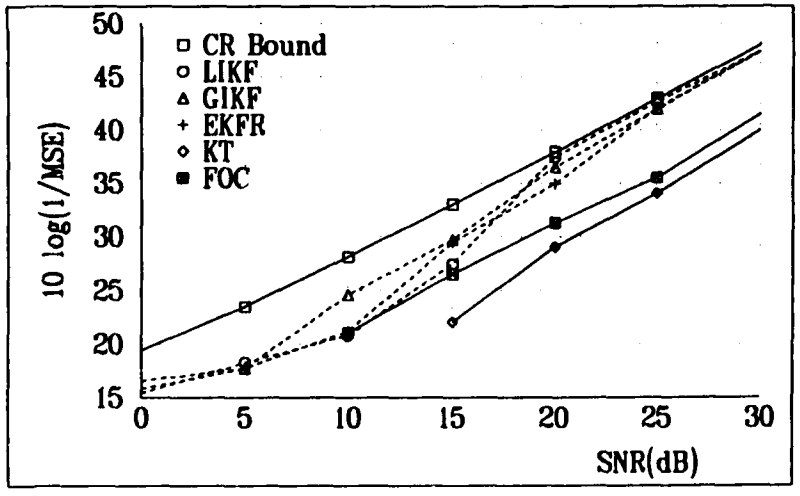


(d) Filter Performance for $\alpha_{k_2} = 0.1$

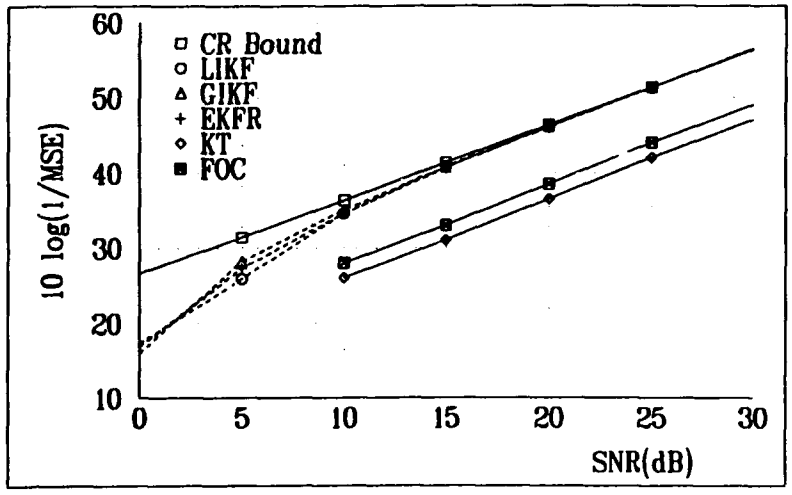
Figure 4.7 Noniterated Filter Performance in White Noise, 4-State Model, $P_0 = 0.04\ I$



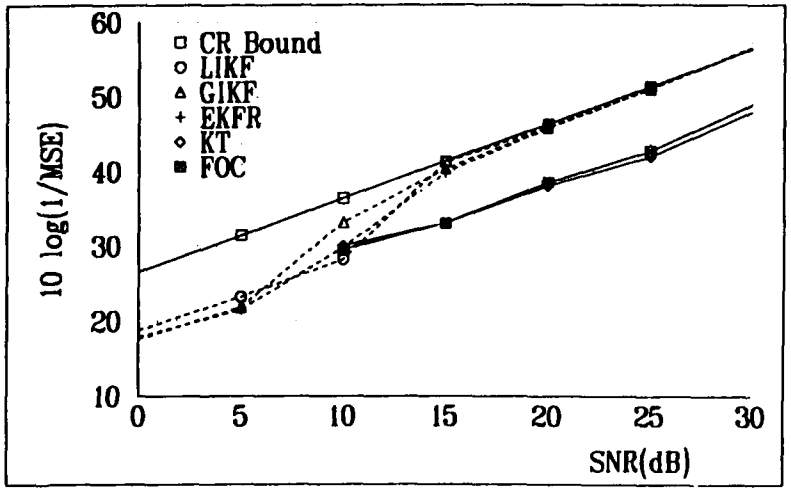
(a) Filter Performance for $\omega_{k_1} = 0.42 * 2\pi$



(b) Filter Performance for $\alpha_{k_1} = 0.2$



(c) Filter Performance for $\omega_{k_2} = 0.52 * 2\pi$



(d) Filter Performance for $\alpha_{k_2} = 0.1$

Figure 4.8 Iterated Filter Performance in White Noise, 4-State Model, $P_0 = 0.04\ I$

performance smoothly degrades as the SNR is decreased. That is, there is no sharp dropoff in performance around the 15 dB point as there is with the least squares approaches. However, there is evidence that there is some kind of threshold effect in the filter performance. In Figure 4.7 the curves representing the estimation error for the two damping coefficients (Figures 4.7(b) and 4.7(d)) change slope between 10 dB and 20 dB SNR. At 10 dB the slope changes back to be roughly parallel to the CR-bound curve. As shown in the Appendix, the performance of the filters has a lower bound represented by the statistics of the initial estimation error. That is, the sample variance of the estimation error cannot be any worse than the initial covariance P_0. This constrains the worst case performance of these nonlinear estimators.

EKF-type algorithms have been known to diverge for poor initial estimates. In some cases these poor estimates lead to poor filter performance due to the approximations made by 1^{st} and 2^{nd} order Taylor series expansions. A test was devised to detect situations with poor final state estimates based on the sample variance of the time series generated by subtracting the estimated measurement, formed from the final state estimates, from the actual measurements. This test is described in Section 4.4. The test works best at high SNR's where the variance of the signal plus noise is significantly better than the variance of the noise alone. The results in Figures 4.7 and 4.8 are those which have passed this test. Figure 4.9 shows the number of runs which passed the test as a function of SNR for each of the six filters for $P_0 = 0.04$. In general, more runs were discarded by the noniterated filters than by the iterated filters. Among the iterated filters there is no consistent better performer. The EKF discarded many more runs than any other filter. The MVF and the GSO performed about the same - significantly better than the EKF, but

worse in general than the iterated filters. The iterated filters not only discarded less runs due to poor performance, but among those runs that were considered valid, the iterated filters gave better overall sample variance.

4.3 Harmonic Retrieval in Colored Noise

Consider the case where the measurement noise is colored by the single pole filter model

$$\mathbf{u}_k = \gamma_{k-1}\mathbf{u}_{k-1} + \mathbf{v}_{k-1} \tag{4.33}$$

where $\mathbf{u}_k$ is the measurement noise, and $\mathbf{v}_k$ is a zero mean white Gaussian noise sequence with $\mathbf{v}_k \sim N(0, R_k)$. The parameter γ_{k-1} is the filter coefficient. To accommodate the colored noise state augmentation or measurement differencing can be applied when γ_{k-1} is known. The complete derivation of the filter equations in colored noise is given in [65]. The plant and measurement models for the colored noise model are given by

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k \\ \mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k) + \mathbf{u}_k \end{aligned} \tag{4.34}$$

Here, $\mathbf{x}_0$, $\mathbf{v}_0$, and $\mathbf{u}_k$ are mutually independent and Gaussian. We have $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, P_0)$, and $\mathbf{u}_0 \sim N(0, U_0)$.

4.3.1 Colored Noise - Known Filter Coefficient

When the filter coefficient γ_{k-1} is known, a set of equivalent "derived" measurements is obtained by subtracting $\gamma_{k-1}\,\mathbf{z}_{k-1}$ from (4.34) to obtain

$$\bar{\mathbf{z}}_k = \mathbf{z}_k - \gamma_{k-1}\mathbf{z}_{k-1} = \mathbf{h}_k(\mathbf{x}_{k-1}) - \gamma_{k-1}\mathbf{h}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{v}_k \tag{4.35}$$

where $\mathbf{x}_k = \mathbf{x}_{k-1}$. The augmented measurement $\bar{\mathbf{z}}_k$ is a nonlinear function of the state with additive white noise.

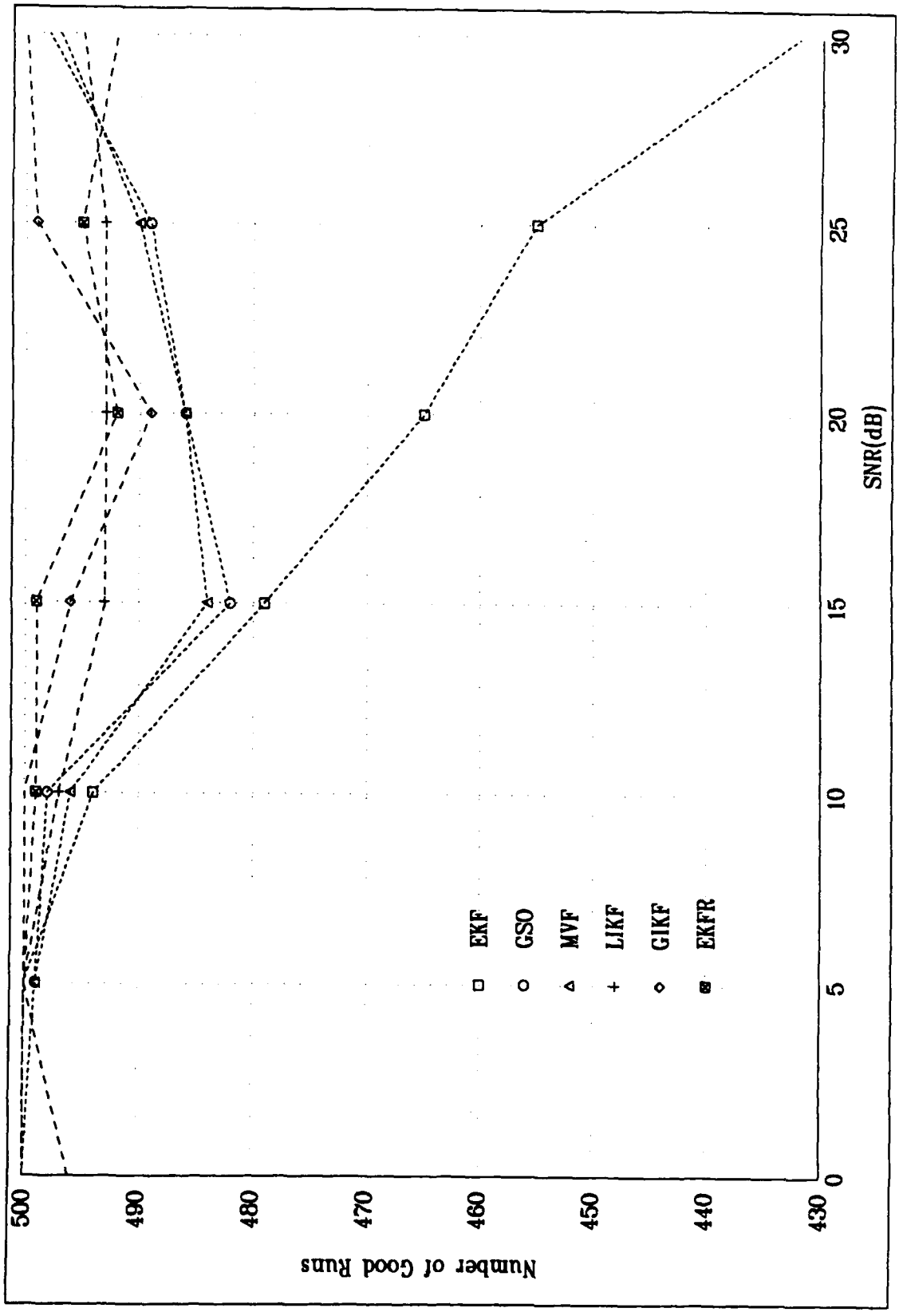


Figure 4.9 Noise Discriminator Results for the Six Nonlinear Filters, 4-State Model, $P_0 = 0.04\ I$

By applying the usual extended Kalman filter linearization of the measurement equation, the filter equation

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k-1|k-1} + K_k \left[\bar{\mathbf{z}}_k - \bar{H}_k \hat{\mathbf{x}}_{k-1|k-1}\right] \tag{4.36}$$

is obtained, where $\bar{H}_k = H_k - \gamma_{k-1} H_{k-1}$ and both H_k and H_{k-1} are evaluated at $\hat{\mathbf{x}}_{k-1|k-1}$. The filter gain and covariance propagation equations are now given by

$$\begin{aligned} K_k &= P_{k|k-1}\bar{H}_k^T(\bar{H}_k P_{k|k-1}\bar{H}_k^T + R_k)^{-1} \\ P_{k|k} &= (I_n - K_k\bar{H}_k)P_{k|k-1}. \end{aligned} \tag{4.37}$$

The initial conditions for this estimator are given by

$$\begin{aligned} E[\mathbf{x}_0] &= \hat{\mathbf{x}}_0 + \mathrm{Var}[\mathbf{x}_0]H_0^T[H_0\mathrm{Var}[\mathbf{x}_0]H_0^T + R_0]^{-1}[\mathbf{z}_0 - H_0\hat{\mathbf{x}}_0] \\ P_0 &= \mathrm{Var}[\mathbf{x}_0] - \mathrm{Var}[\mathbf{x}_0]H_0^T[H_0\mathrm{Var}[\mathbf{x}_0]H_0^T + R_0]^{-1}H_0\mathrm{Var}[\mathbf{x}_0] \end{aligned} \tag{4.38}$$

4.3.2 Colored Noise - Unknown Filter Coefficient

If the filter coefficient γ_k is unknown, the state vector can be augmented to include the unknown parameter γ_k. The augmented state vector is defined as

$$\bar{\mathbf{x}}_k \equiv \begin{bmatrix} \mathbf{x}_k \\ \cdots \\ \gamma_k \end{bmatrix}. \tag{4.39}$$

H_k is then defined by

$$H_k = \left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k)}{\partial \bar{\mathbf{x}}_k}\right|_{\bar{\mathbf{x}}_k = \hat{\bar{\mathbf{x}}}_{k-1|k-1}}. \tag{4.40}$$

Let $\mathbf{g}_k(\bar{\mathbf{x}}_k) = \gamma_k \mathbf{h}_k(\mathbf{x}_k)$, such that

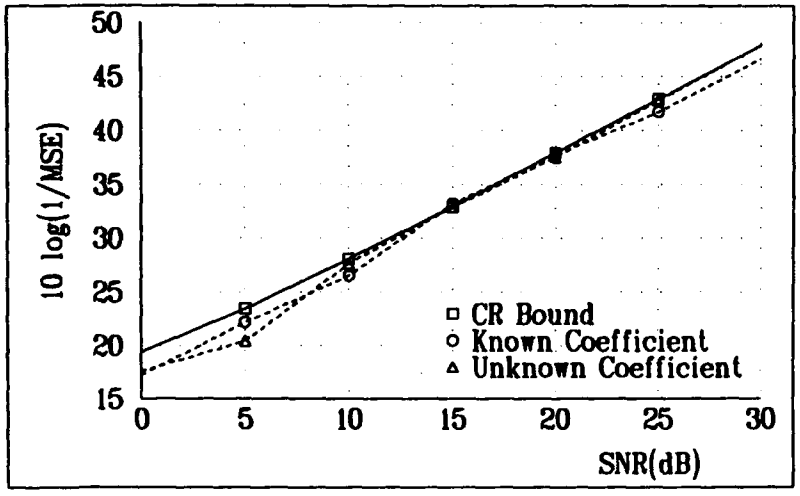
$$G_k = \left.\frac{\partial \mathbf{g}_k(\bar{\mathbf{x}}_k)}{\partial \bar{\mathbf{x}}_k}\right|_{\bar{\mathbf{x}}_k = \hat{\bar{\mathbf{x}}}_{k-1|k-1}}. \tag{4.41}$$

If $\bar{H}_k = H_k - G_{k-1}$ then the filter equations are given by

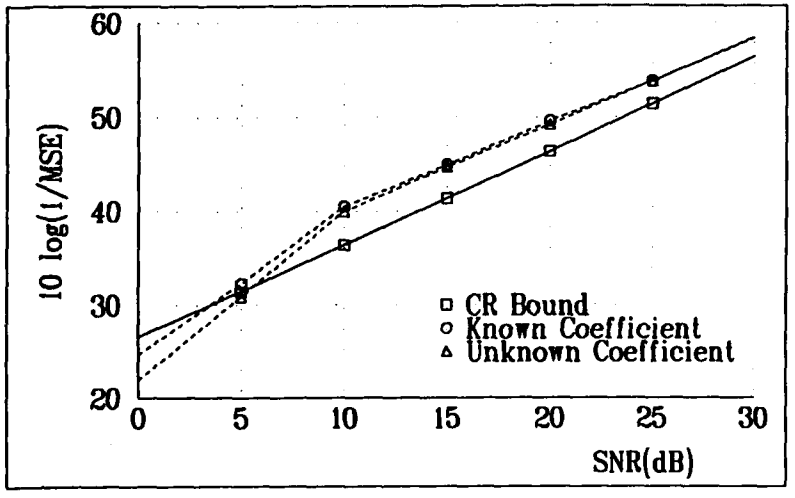
$$\begin{aligned}
\hat{\bar{\mathbf{x}}}_{k|k} &= \hat{\bar{\mathbf{x}}}_{k-1|k-1} + K_k[\bar{\mathbf{z}}_k - \bar{H}_k \hat{\bar{\mathbf{x}}}_{k-1|k-1}] \\
\bar{P}_{k|k} &= (I_n - K_k \bar{H}_k)\bar{P}_{k|k-1} \\
K_k &= \bar{P}_{k|k-1}\bar{H}_k^T(\bar{H}_k \bar{P}_{k|k-1}\bar{H}_k^T + R_k)^{-1}.
\end{aligned} \tag{4.42}$$

4.3.3 Experimental Results for Estimation in Colored Noise

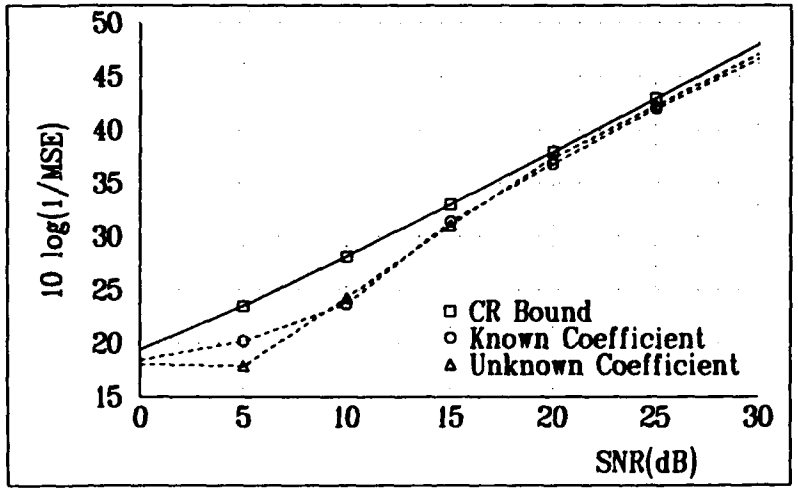
The EKF was used to estimate the model parameters of (4.6) with colored noise given by model (4.33). The simulation results from 500 Monte Carlo trials are presented in Figure 4.10 for initial uncertainty $P_0 = 0.04I$. The filter coefficient is $\gamma_{k-1} = 0.8$. Results are shown in this figure for case of known and unknown γ_k. The solid lines in the four plots show the white noise CR bound, which is not applicable in this case, but is included as a reference. Figure 4.10(d) shows the EKF performance when the filter coefficient is unknown and estimated along with the model parameters. Due to the frequency response of the colored noise filter one would expect the estimates for the first sinusoid to be slightly worse relative to the white noise CR bound than for the second sinusoid. That is, the gain due to the colored noise filter at the frequency of the first sinusoid is higher than that at the second sinusoid. This is verified by Figure 4.10. This figure also shows that the estimation results of the model parameters when the coefficient γ_{k-1} is unknown and is estimated on-line, are only slightly inferior to the results when γ_{k-1} is known.



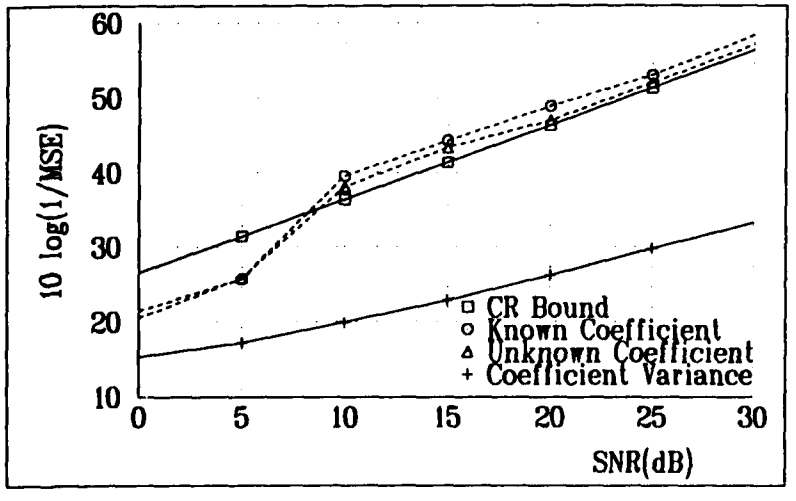
(a) Filter Performance for $\omega_{k_1} = 0.42 * 2\pi$



(c) Filter Performance for $\omega_{k_2} = 0.52 * 2\pi$



(b) Filter Performance for $\alpha_{k_1} = 0.2$



(d) Filter Performance for $\alpha_{k_2} = 0.1$, $\gamma_k = 0.8$

Figure 4.10 EKF Performance in Colored Noise with Known and Unknown Noise Coefficient, 4-State Model, $P_0 = 0.04\ I$

4.4 Estimation of the Measurement Covariance

In many situations of practical interest, the measurement noise R is unknown. In the following development it is assumed that the measurement noise is stationary over all of the measurements and that the states are constant. An initial estimate of the noise R_0 is used to initialize the filter. Given the measurements $Z_k = (\mathbf{z}_1; \mathbf{z}_2; \cdots; \mathbf{z}_k)$, R_k can be estimated by taking the sample variance of the innovations error at each iteration of the filter. Let S_k be defined by

$$\begin{aligned} S_k &= E[\tilde{\mathbf{z}}_k \tilde{\mathbf{z}}_k^T] \\ &= H_k P_{k|k-1} H_k^T + R_k. \end{aligned} \tag{4.43}$$

Given H_k and $P_{k|k-1}$ an estimate of S_k is required in order to obtain an estimate for the measurement noise R_k. Let Θ_k be the matrix of estimates

$$\Theta_k = \left[\mathbf{h}_1(\mathbf{x}_1)|_{\mathbf{x}_1=\hat{\mathbf{x}}_{k|k-1}}; \mathbf{h}_2(\mathbf{x}_2)|_{\mathbf{x}_2=\hat{\mathbf{x}}_{k|k-1}}; \cdots, ; \mathbf{h}_k(\mathbf{x}_k)|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}}\right]. \tag{4.44}$$

Θ_k and Z_k are m by K matrices where m is the number of elements in the measurement vector, and K represents the total number of time intervals. Let the innovations error matrix be defined by

$$E_k = Z_k - \Theta_k, \tag{4.45}$$

and let $\mathbf{e}_{k_i}$ be the i^{th} row of E_k. Assuming that the innovations are zero mean, an estimate of the ij^{th} element of $\hat{S}_k$ can be obtained using the sample variance

$$\hat{s}_{k_{ij}} = \frac{1}{K-1} \mathbf{e}_{k_i} \mathbf{e}_{k_i}^T. \tag{4.46}$$

The measurement variance can then be estimated using

$$\hat{R}_k = \hat{S}_k - H_k P_{k|k-1} H_k^T. \tag{4.47}$$

In the situation where the $P_{k|k-1}$ is equal to zero, that is when the estimates are perfect, the parameter $\hat{s}_{k_{ij}}$ gives the sample variance of the measurement noise. The statistic $(K-1)\hat{s}_{k_{ij}}$ is chi-square distributed with K degrees of freedom. The chi-square distribution, represented by the sample variance statistic, has a 99.9% probability of being less than or equal to $2\sigma^2$ for $K = 50$, where σ^2 is the variance of the measurement noise. The statistic $\hat{s}_{k_{ij}}$ can be used as a way to detect poor state estimates. A reasonable criterion is to reject the estimate as probably bad if $\hat{s}_{k_{ij}}$ is greater than $2\sigma^2$. This criterion works very well at high SNR's where the variance of signal plus noise is significantly different from the variance of noise alone. However, at low SNR's the variance of signal plus noise can be close to the variance of the noise alone and this test does not work as well.

4.4.1 Experimental Results - Unknown Noise Statistics

Experimental results of the extended Kalman filter with unknown noise covariance are shown in Figure 4.11. This figure gives two sets of curves. One set shows the filter performance with one pass through the data. During this pass the noise statistics are estimated at each iteration using equation (4.35). The initial noise covariance was estimated $\hat{\sigma}^2 = 0.05$, corresponding to 10 dB SNR, for all points on this curve. The second curve shows the results of processing the data with two passes. During the first pass the noise is estimated as in the first case. During the second pass the noise covariance estimate is held constant at the final value from the first pass and the data is filtered using the normal extended Kalman

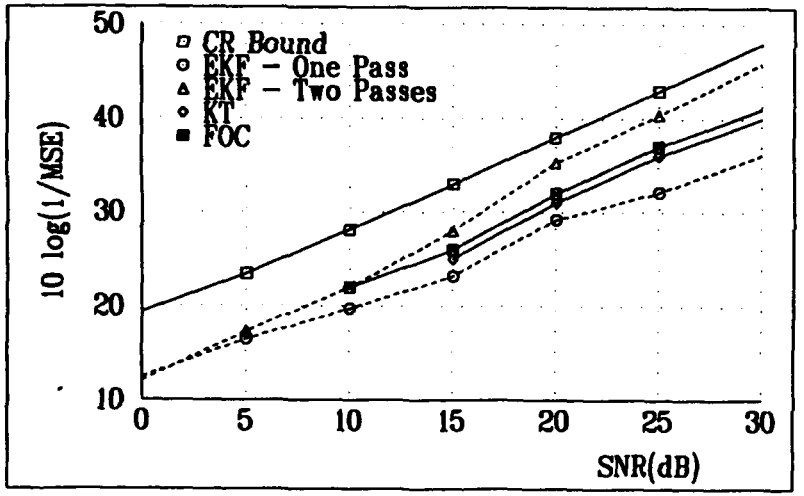
filter relations. For this second pass the initial state covariance is reset to 0.04. The initial state estimate for the second pass was equal to the final state estimate from the first pass. The results show that the two-pass filter performs significantly better than the single-pass filter, particularly at high SNR's. At low SNR's the two filters result in about the same performance. Comparing Figure 4.11 to Figure 4.7 the two-pass filter with initially unknown noise statistics results in about the same performance as the (single-pass) extended Kalman filter with known noise statistics.

4.4.2 Experimental Results - Single Sinusoid

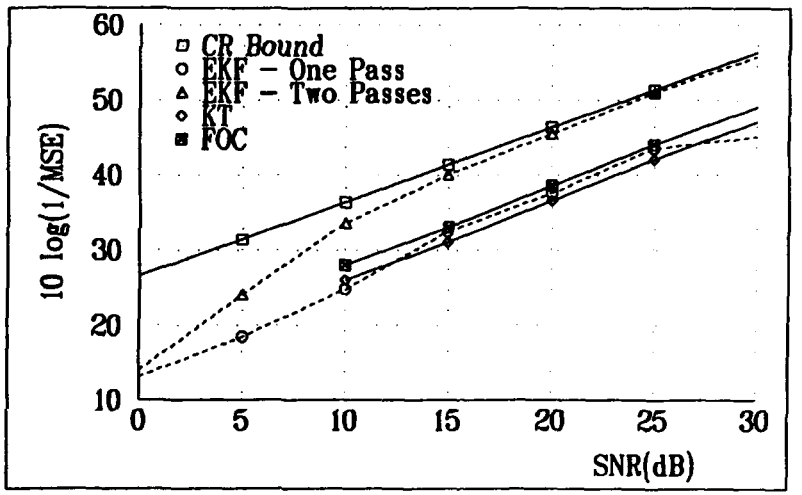
A study was also performed on a single sinusoid model. This model was formed by letting $P = 1, c_{k_1} = 1, \alpha_{k_1} = 0.12, \omega_{k_1} = 0.22*2\pi$, and $\theta_{k_1} = 0$ in equation (4.6). The unknown random parameters in this system were α_{k_1} and ω_{k_1}. The *sample variances* for all *six* of the filters is given in Figure 4.12 for initial estimation error variance $P_0 = 0.09$ for both state variables. Figure 4.13 presents the sample variance for the model with initial estimation error uniformly distributed between 0 and 2π for the frequency, corresponding to no a *priori* information, and with initial estimation error uniformly distributed between 0 and 1 for the damping coefficient. The results are similar to those in Figure (4.12) for high SNR, especially for the LIKF. However, for low SNR the performance is significantly worst in (4.13) than in (4.12). These results satisfy intuition in that whenever there are less parameters to be estimated the filter can sustain larger initial estimation error.

4.5 Conclusion

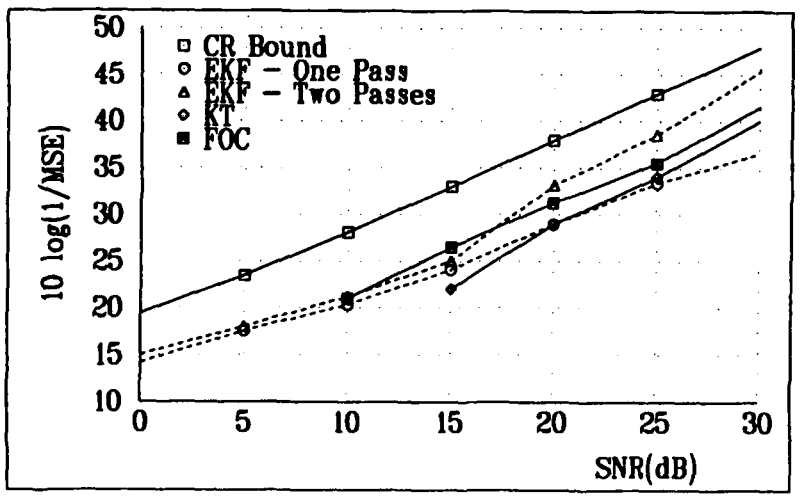
Methods based on nonlinear recursive filters for estimating the parameters of exponentially damped sinusoids in white and colored noise have been described. Filter equations have been developed for time varying systems in white and colored



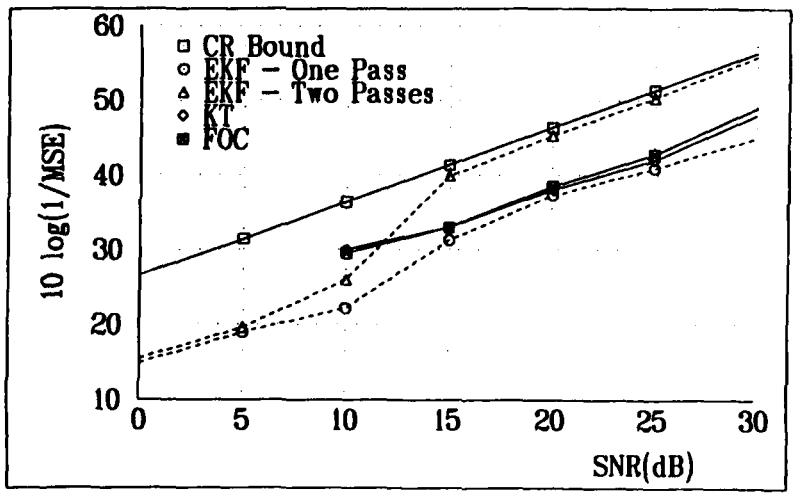
(a) Filter Performance for $\omega_{k_1} = 0.42 * 2\pi$



(c) Filter Performance for $\omega_{k_2} = 0.52 * 2\pi$

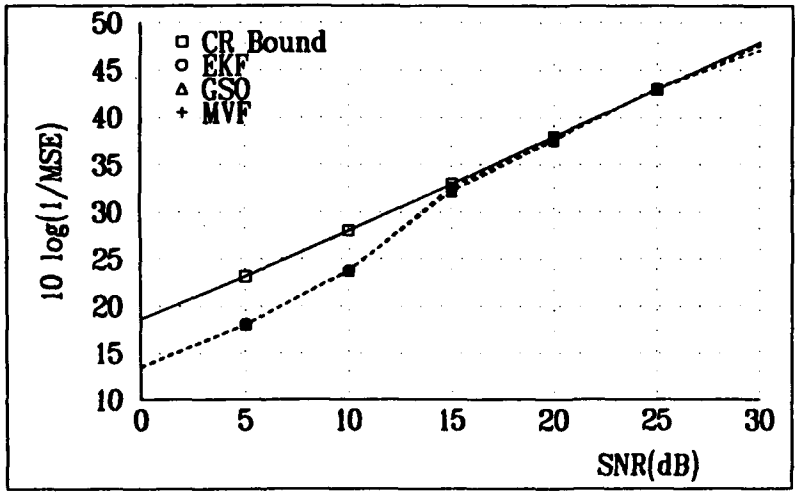


(b) Filter Performance for $\alpha_{k_1} = 0.2$

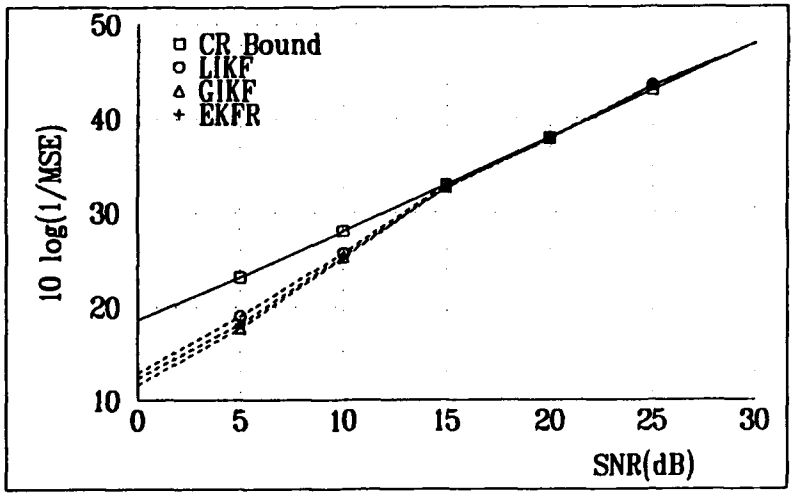


(d) Filter Performance for $\alpha_{k_2} = 0.1$

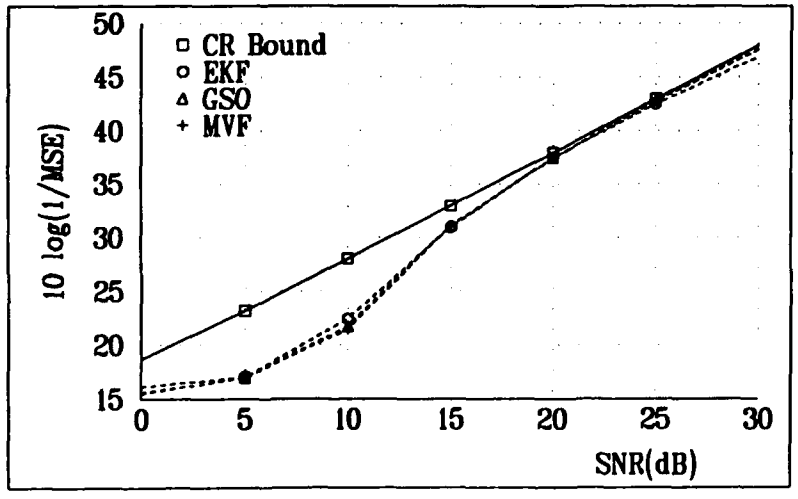
Figure 4.11 EKF Performance in White Noise with Unknown Noise Covariance, 4-State Model, $P_0 = 0.04\ I$



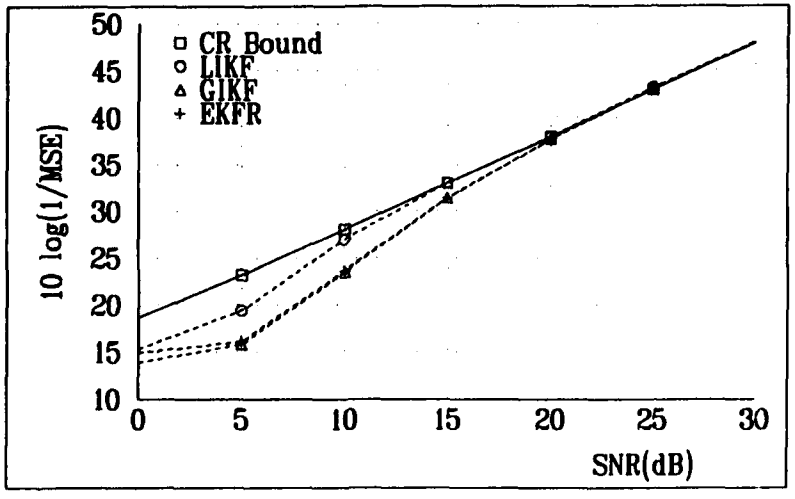
(a) Filter Performance for $f_k = 0.52$



(c) Filter Performance for $f_k = 0.52$

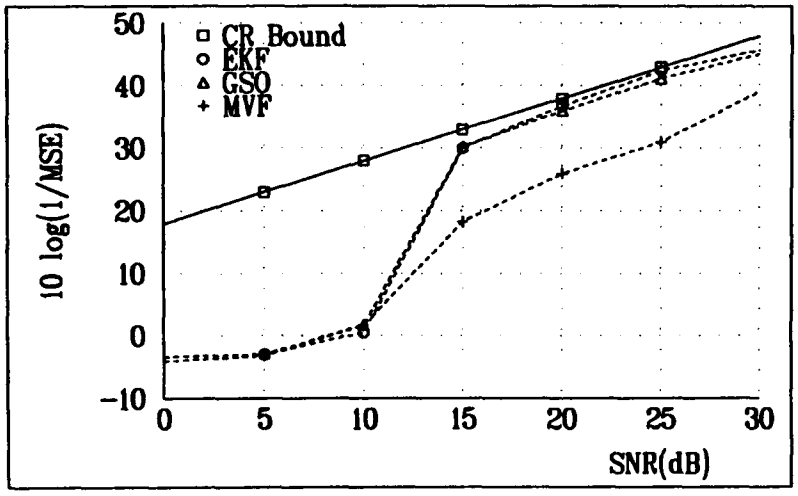


(b) Filter Performance for $\alpha_k = 0.2$

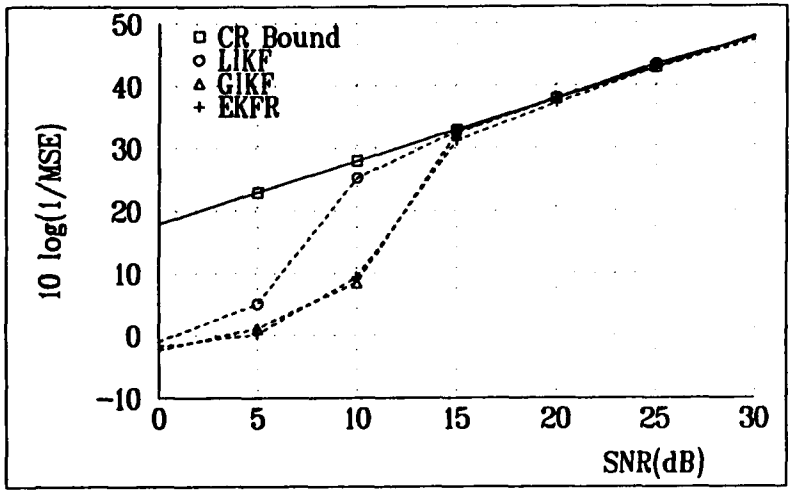


(d) Filter Performance for $\alpha_k = 0.2$

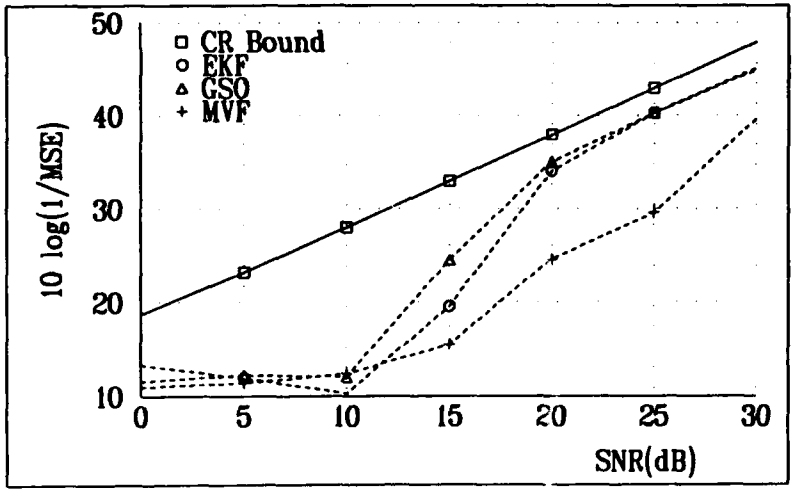
Figure 4.12 Performance of All Six Nonlinear Filters, 2-State Model, $P_0 = 0.09\ I$



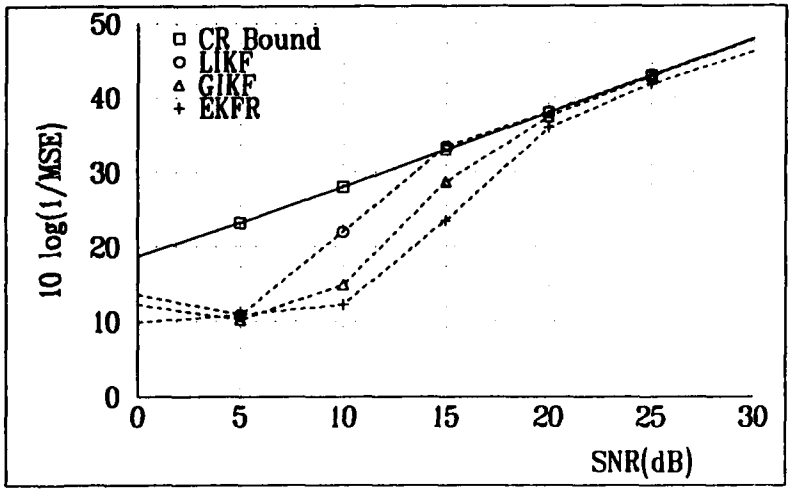
(a) Filter Performance for $f_k = 0.52$



(c) Filter Performance for $f_k = 0.52$



(b) Filter Performance for $\alpha_k = 0.2$



(d) Filter Performance for $\alpha_k = 0.2$

Figure 4.13 Performance of All Six Nonlinear Filters, 2-State Model, Large P_0

noise with known and unknown noise covariances. Simulation results for the problem of estimating the parameters of two exponentially damped sinusoids show that the nonlinear filtering techniques described perform very close to the Cramer-Rao bound, even at low SNR's, for relatively small initial estimation errors. For larger values of the initial uncertainty on the model parameters the iterated forms of the extended Kalman filter give better performance than the noniterated forms. Among the noniterated forms the Gaussian second order filter and the minimum variance filter give comparable performance, and both perform better than the extended Kalman filter, particularly at high SNR's. In addition these two high order filters are generally more stable as evidenced by the number of final state estimates that passed the noise discriminator test. The extended Kalman filter has been shown to give good results in colored noise with known and unknown noise filter parameter. A technique has also been developed for on-line estimation of the measurement noise covariance.

In summary the following general observations can be drawn about the performance of nonlinear filters for harmonic retrieval:

* The nonlinear filters incorporate *a priori* knowledge about the state. The KT method has no inherent capability to use *a priori* knowledge.

* As with the KT method the nonlinear filter method approaches the CR bound at high SNR's. However, the performance of the nonlinear filters does not degrade sharply in the range of 10 to 15 dB as the performance of the KT method does. Worst case performance of the nonlinear filters is bounded by the initial error covariance.

* The nonlinear filters can estimate parameters in colored noise. The KT

method is not designed to operate in colored noise.

* The filters converge relatively fast, making the nonlinear filters suitable for short data lengths.

* The nonlinear filters are recursive in nature, thereby providing adaptability to time varying parameters.

* The primary disadvantage of the nonlinear filter methods is that they may require good initial estimates to converge to a valid solution. The KT method does not share this problem. However, the multi-filter resolution approach, which is defined in Chapter 5 and implemented in Chapter 7 for time delay estimation, can be used to accommodate poor initial conditions by partitioning them into smaller intervals of uncertainty and applying joint detection/estimation techniques for resolving ambiguity.

* Another interesting approach to performing harmonic retrieval in with large initial estimation error would be to use the KT method to initialize the nonlinear filter. The nonlinear filter would then be to refine the estimates.

Chapter 5

Joint Detection/Estimation

This chapter presents a procedure for combining detection and estimation theory. This procedure is used in subsequent chapters for selected signal processing problems. Two general applications of joint detection/estimation theory are addressed. In the first application the nonlinear measurement model is constant, but the initial estimation error is large enough such that the approximations made by nonlinear estimators such as the extended Kalman filter may lead to very poor performance. In this case the a *priori* pdf can be partitioned into M smaller subregions. Each subregion is associated with a hypothesis and the problem is treated as an M-ary hypothesis joint detection/estimation problem. In the second application the measurement model is unknown. Again, several hypotheses are proposed. Each hypothesis is associated with a different model. For each model the state variables are estimated on line. This operation is performed concurrent with estimation the a *posteriori* probability of each hypothesis. The states are not constrained to be common among the models. A third application involves a combination of the other two applications. This chapter presents the general technique for applying joint detection/estimation theory to these applications.

The traditional estimation theory approach to solving parameter estimation problems involves starting with initial estimates of the state variables and refining these estimates by filtering the measurement data. The performance of the filter is governed by the statistics of the process and measurement noises, and by the process and measurement models. Filters such as the Kalman filters may exhibit unstable behavior. For nonlinear process or measurement models the approximations result-

ing from truncated Taylor series expansions used in the implementation of the EKF may also lead to poor performance. It is well known that under certain conditions the extended Kalman filter may give poor results due to the approximation made with the first order Taylor series expansion. This is particularly evident when the initial estimation error is large. This problem is discussed in the context of the harmonic retrieval problem in Chapter 4.

M-ary detection is used in a number of practical signal processing problems. One example of this type of detection is ambiguity function processing for radar/sonar signal processing. This process involves the convolution or matched filtering of a received signal with a number of signal replicas. Each replica has a different estimate of the unknown states (e.g. amplitude, delay, Doppler shift). The replica that results in the largest value of the ambiguity function is chosen is used to determine the state estimates. Autoregressive model order selection may also be treated using M-ary detection theory where the M hypotheses correspond to all of the combinations of a discrete set of frequencies. These two problems are addressed in Chapters 6-8 of this thesis using the joint detection/estimation approach which is presented in this chapter.

The joint detection/estimation (JD/E) approach combines M-ary detection with estimation. The distinction of JD/E over pure M-ary detection is that the estimates are refined for each hypothesis. JD/E may permit the use of a smaller number of hypotheses than detection only since the hypotheses are continuously refined through estimation, which is performed concurrently with the hypothesis testing. Alternatively, one may use the same number of hypotheses as the detection only problem, but by refining the estimates through filtering, better estimates may result. A major reason for using JD/E in nonlinear filtering problems is to reduce

the effects of large initial estimation error on the truncated Taylor series expansion used in the EKF.

In this chapter a recursive technique for joint detection/estimation is developed based on nonlinear state and measurement models. The objective is to develop a procedure that results in an optimal minimum variance estimate for the state variables for each hypothesis and, given this optimal estimate, to select the proper hypothesis which most closely matches the measurement data.

The development of the joint detection/estimation is based on the segmentation of the unknowns into a state vector $\mathbf{x}_k$ and a parameter vector θ. Let $\theta \in \Theta$ designate the parameter vector that describes the different models that may have generated the measurements. Each model is identified with a specific hypothesis, and corresponds to a unique θ, $\theta \in \Theta$. The set Θ is assumed to be countable (in our application also finite). In addition, the parameter vector θ is assumed time invariant. The development of the joint detection/estimation presented here follows a similar procedure to that presented by Fredriksen *et al* [58]. However, these authors combined the state vector $\mathbf{x}_k$ and the parameter vector θ together into one state vector that was used in the estimation process. In addition, it was required that all of the variables in the augmented state vector be energy variables. In the development that follows a distinction is made between the state vector $\mathbf{x}_k$ and the parameters θ. The state variables are the same for each hypothesis, while the vector θ is used to distinguish between the various hypotheses. There is no restriction on the state vector $\mathbf{x}_k$. The parameter vector θ is assumed to be time invariant. Under hypothesis H_θ the discrete time measurements are modeled according to

$$H_\theta \quad : \quad \mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \theta) + \mathbf{v}_{k,\theta} \tag{5.1}$$

where the state $\mathbf{x}_k$ is common for all $\theta \in \Theta$, and satisfies the discrete time process equation

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \theta) + \mathbf{w}_{k-1,\theta} \tag{5.2}$$

with initial state estimate $\hat{\mathbf{x}}_{0|0,\theta}$, and initial state covariance $P_{0|0,\theta}$. The initial state estimate, the measurement noise, and the process noise are uncorrelated up to the moment order required by the implemented filter (e.g EKF or EHOF). The process and measurement noise are zero mean and distributed with covariances $E[\mathbf{w}_{k,\theta}\mathbf{w}_{k,\theta}^T] = Q_{k,\theta}$, and $E[\mathbf{v}_{k,\theta}\mathbf{v}_{k,\theta}^T] = R_{k,\theta}$.

The model structure in (5.1) and (5.2) is similar to the traditional state and observation models with the exception that an unknown time invariant parameter vector θ is used to distinguish between the various hypotheses. This model structure is designed to accomodate a large class of process and measurement models. The only restriction is that the state vector is the same for each hypothesis. Note, however, that not all of the elements of the state vector $\mathbf{x}_k$ are required to be estimated under each hypothesis.

It is assumed that θ has known probability density function $p(\theta)$. The vector θ is required to contain coefficients of a sufficient set of energy parameters (e.g. amplitude, time duration) such that the null hypothesis is indicated whenever this set of parameters is equal to zero. Although it is not mandatory that all of the parameters in θ be zero to indicate the null hypothesis, it simplifies the discussion of the method. Thus, under H_0, $\theta_0 = 0$, and $\mathbf{h}_k(\mathbf{x}_k, \theta_0) = \mathbf{h}_k(\mathbf{x}_k, 0) = 0$.

5.1 A Bayes Test for Joint Detection/Estimation

The Bayesian approach for optimum detection involves the minimization of

the average decision cost over all possible decisions under all possible hypotheses. Let $\mathbf{Z}_k = \{\mathbf{z}_1, \mathbf{z}_2, \cdots \mathbf{z}_k\}$, be the collection of measurements that are functions of the time varying state variables $\mathbf{x}_k$, and the unknown time invariant parameter vector θ. The total average cost, or Bayes Risk for joint detection/estimation, of making the decisions D_j associated with hypotheses H_j, $j = 0, \cdots, M$ can be expresses as

$$\Re_{D\&E} = \sum_{j=0}^{M} \int_{R_j} \int \int C_j(\mathbf{x}_k) p(\mathbf{z}_k, \mathbf{x}_k, \theta | \mathbf{Z}_{k-1}) d\theta d\mathbf{x}_k d\mathbf{z}_k \tag{5.3}$$

where $C_j(\mathbf{x}_k)$ is the cost of making decision j, and $p(\mathbf{z}_k, \mathbf{x}_k, \theta | \mathbf{Z}_{k-1})$ is the density function of all of the random parameters in the system, given the measurements up to $k-1$. The goal is to find the estimate $\hat{\mathbf{x}}_{k|k}$ that minimizes the total average cost. By conditioning the decision probability on the past measurements a recursive technique for joint detection estimation can be developed.

If the cost $C_j(.)$ is not a function of the state $\mathbf{x}_k$ (i.e, the detection only case [75]), then C_j can be moved outside of the integral and the risk becomes

$$\Re = \sum_{j=0}^{M} C_j P(D_j | \mathbf{Z}_{k-1}) \tag{5.4}$$

where $P(D_j | \mathbf{Z}_{k-1})$ is given by

$$P(D_j | \mathbf{Z}_{k-1}) = \int_{R_j} \int \int p(\mathbf{z}_k, \mathbf{x}_k, \theta | \mathbf{Z}_{k-1}) d\theta d\mathbf{x}_k d\mathbf{z}_k \tag{5.5}$$

By applying Bayes rule to the joint density function in (5.3), the Bayes risk for joint detection/estimation can be expressed as

$$\Re_{D\&E} = \sum_{j=0}^{M} \int_{R_j} \int \int C_j(\mathbf{x}_k) p(\mathbf{z}_k | \mathbf{x}_k, \theta) p(\mathbf{x}_k | \mathbf{Z}_{k-1}, \theta) p(\theta | \mathbf{Z}_{k-1}) d\theta d\mathbf{x}_k d\mathbf{z}_k \tag{5.6}$$

where $p(\mathbf{z}_k | \mathbf{x}_k, \theta) = p(\mathbf{z}_k | \mathbf{Z}_{k-1}, \mathbf{x}_k, \theta)$ because of the Markov noise process in (5.1).

That is, it is easily seen from (5.1) that if the $\mathbf{x}_k$, and θ are known, then the density function for $\mathbf{z}_k$ is a function of the measurement noise $\mathbf{v}_k$ only. Let the parameter space spanned by θ be discrete such that it can be characterized by a finite number of quantized points. Then the a *priori* probability density of θ given the measurements $\mathbf{Z}_{k-1}$ can be expressed as

$$p(\theta|\mathbf{Z}_{k-1}) = \sum_{i=0}^{M} P(\theta_i|\mathbf{Z}_{k-1}) p_i(\theta|\mathbf{Z}_{k-1}) = \sum_{i=0}^{M} P(\theta_i|\mathbf{Z}_{k-1}) \delta(\theta - \theta_i) \tag{5.7}$$

where $P(\theta_i|\mathbf{Z}_{k-1})$ is the a *priori* probability of hypothesis H_i given the measurements $\mathbf{Z}_{k-1}$. Thus, $P(\theta_i|\mathbf{Z}_{k-1})$ can be used in place of the more conventional $P(H_i|\mathbf{Z}_{k-1})$ to demonstrate the explicit dependence of each hypothesis on the parameter θ_i. The hypothesis H_i then corresponds to

$$H_i \quad : \quad \mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \theta_i) + v_{k,\theta_i} \tag{5.8}$$

Given the measurements $\mathbf{Z}_{k-1}$, the cost associated with hypothesis j [75, pp. 140-141] can be expressed as

$$\begin{aligned} C_j(\mathbf{x}_k) &= \sum_{i=0}^{M} C_{ji}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k) \frac{P(\theta_i|\mathbf{Z}_{k-1}) p_i(\theta|\mathbf{Z}_{k-1})}{p(\theta|\mathbf{Z}_{k-1})} \\ &= \sum_{i=0}^{M} C_{ji}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k) \frac{P(\theta_i|\mathbf{Z}_{k-1}) \delta(\theta - \theta_i)}{p(\theta|\mathbf{Z}_{k-1})} \end{aligned} \tag{5.9}$$

where $C_{ji}(.)$ is the cost of deciding H_j given H_i is true. Substituting (5.9) into (5.6), the Bayes risk becomes

$$\Re_{D\&E} = \sum_{j=0}^{M} \sum_{i=0}^{M} \int_{R_j} \int P(\theta_i|\mathbf{Z}_{k-1}) C_{ji}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k) p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k d\mathbf{z}_k \tag{5.10}$$

The risk associated with deciding hypothesis H_0, the null hypothesis, is found by

evaluating (5.10) for $j = 0$. Noting that if the decision regions are disjoint, then $R_0 = R - \sum_{j=1}^{M} R_j$, and this risk becomes

$$\begin{aligned} C_0(\mathbf{x}_k)P(D_0|\mathbf{Z}_{k-1}) &= \sum_{i=0}^{M} \int_{R_0} \int P(\theta_i|\mathbf{Z}_{k-1}) C_{0i}(\mathbf{x}_k) p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k d\mathbf{z}_k \\ &= \sum_{i=0}^{M} \int_{R} \int P(\theta_i|\mathbf{Z}_{k-1}) C_{0i}(\mathbf{x}_k) p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k d\mathbf{z}_k \\ &\quad - \sum_{i=0}^{M} \int_{\sum_{j=1}^{M} R_j} \int P(\theta_i|\mathbf{Z}_{k-1}) C_{0i}(\mathbf{x}_k) \\ &\qquad \times p(\mathbf{z}_k|\mathbf{x}_k, \theta) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k d\mathbf{z}_k \end{aligned} \tag{5.11}$$

where the explicit dependence of $C_{0i}(\mathbf{x}_k)$ on $\hat{\mathbf{x}}_{k|k}$ has been removed because an estimate is not required whenever hypothesis H_0 is decided. Using the result (5.11) in (5.10) the total risk now becomes

$$\begin{aligned} \Re_{D\&E} &= \sum_{i=0}^{M} \int_{R} \int P(\theta_i|\mathbf{Z}_{k-1}) C_{0i}(\mathbf{x}_k) p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k d\mathbf{z}_k \\ &\quad + \sum_{j=1}^{M} \sum_{i=0}^{M} \int_{R_j} \int P(\theta_i|\mathbf{Z}_{k-1}) \left[C_{ji}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k) - C_{0i}(\mathbf{x}_k) \right] \\ &\qquad \times p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k d\mathbf{z}_k \end{aligned} \tag{5.12}$$

Under the null hypothesis H_0, the cost is not a function of $\mathbf{x}_k$. Thus, $C_{j0}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k) = C_{j0}(\hat{\mathbf{x}}_{k|k})$. In addition, C_{00} is neither a function of $\hat{\mathbf{x}}_{k|k}$ nor of $\mathbf{x}_k$. Furthermore, under the null hypothesis, $p(\mathbf{z}_k|\mathbf{x}_k, \theta_0) = p(\mathbf{z}_k|\theta_0)$. That is, the density function for the measurements is not a function of the state variables. This

leads to

$$\int P(\theta_0|\mathbf{Z}_{k-1})\,[C_{j0}(\hat{\mathbf{x}}_{k|k}) - C_{00}]\,p(\mathbf{z}_k|\mathbf{x}_k,\theta_0)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_0)d\mathbf{x}_k = P(\theta_0|\mathbf{Z}_{k-1})\,[C_{j0}(\hat{\mathbf{x}}_{k|k}) - C_{00}]\,p(\mathbf{z}_k|\theta_0) \tag{5.13}$$

By defining the likelihood ratio

$$L_{ji}(\mathbf{z}_k) \equiv \frac{P(\theta_i|\mathbf{Z}_{k-1})\,\int[C_{0i}(\mathbf{x}_k) - C_{ji}(\hat{\mathbf{x}}_{k|k},\mathbf{x}_k)]\,p(\mathbf{z}_k|\mathbf{x}_k,\theta_i)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_i)d\mathbf{x}_k}{P(\theta_0|\mathbf{Z}_{k-1})\,[C_{j0}(\hat{\mathbf{x}}_{k|k}) - C_{00}]\,p(\mathbf{z}_k|\theta_0)}, \tag{5.14}$$

the Bayesian risk can now be expressed as

$$\begin{aligned}\Re_{D\&E} = &\sum_{i=0}^{M}\int_R\int P(\theta_i|\mathbf{Z}_{k-1})C_{0i}(\mathbf{x}_k)p(\mathbf{z}_k|\mathbf{x}_k,\theta_i)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_i)d\mathbf{x}_k d\mathbf{z}_k \\ &+\sum_{j=1}^{M}\int_{R_j} P(\theta_0|\mathbf{Z}_{k-1})\,[C_{j0}(\hat{\mathbf{x}}_{k|k}) - C_{00}]\,p(\mathbf{z}_k|\theta_0)\left[1 - \sum_{i=1}^{M} L_{ji}(\mathbf{z}_k)\right] d\mathbf{z}_k\end{aligned} \tag{5.15}$$

The first term on the RHS of (5.15) is constant. Hence, it does not contribute to the selection of the decision boundaries. It is assumed that the cost making a wrong decision $C_{ji}(\hat{\mathbf{x}}_{k|k},\mathbf{x}_k)$, $j \neq i$, is greater than the cost of making a correct decision $C_{ii}(\hat{\mathbf{x}}_{k|k},\mathbf{x}_k)$. Since all probabilities and density functions are positive or zero, $P(\theta_0|\mathbf{Z}_{k-1})\,[C_{j0}(\hat{\mathbf{x}}_{k|k}) - C_{00}]\,p(\mathbf{z}_k|\theta_0) \geq 0$. Thus, the decision is made in favor of hypothesis H_j based on the selection

$$j = \text{argmin}_j \left\{[C_{j0}(\hat{\mathbf{x}}_{k|k}) - C_{00}]\left[1 - \sum_{i=1}^{M} L_{ji}(\mathbf{z}_k)\right]\right\} \tag{5.16}$$

This decision rule is based on the fact that the estimate $\hat{\mathbf{x}}_{k|k}$ is optimum.

The second part of the joint detection/estimation procedure is to find the optimum estimate. The optimum estimator, given that decision D_j was decided

after the first stage, will be denoted $\hat{\mathbf{x}}^*_{k|k}$ and is determined by finding the value that minimizes the total average cost in decision region R_j. That is, $\hat{\mathbf{x}}^*_{k|k}$ is determined from the condition

$$\min_{\hat{\mathbf{x}}_{k|k}} \int_{R_j} P(\theta_0|\mathbf{Z}_{k-1}) \left[C_{j0}(\hat{\mathbf{x}}_{k|k}) - C_{00}\right] p(\mathbf{z}_k|\theta_0) \left[1 - \sum_{i=1}^{M} L_{ji}(\mathbf{z}_k)\right] d\mathbf{z}_k. \quad (5.17)$$

The exact form of the optimum estimator is a function of the type of cost function $C_{ji}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k)$.

5.1.1 Quadratic Cost Function

Consider the case where the cost is a quadratic function of the estimation error. The quadratic cost is expressed as

$$C_{ji}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k) = b_{ji} + c_{ji}\left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right]^T \left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right] \quad (5.18)$$

The cost includes the term b_{ji} which represents the conventional cost associated with the problem of detection only, and the cost c_{ji} which accounts for the estimation error. The cost of deciding hypothesis D_j given hypothesis H_0 is related to the estimate only and

$$\begin{aligned} C_{00} &= c_{00} \\ C_{0i}(\mathbf{x}_k) &= b_{0i} + c_{0i}\mathbf{x}_k^T\mathbf{x}_k \qquad i \neq 0 \\ C_{j0}(\hat{\mathbf{x}}_{k|k}) &= b_{j0} + c_{j0}\hat{\mathbf{x}}_{k|k}^T\hat{\mathbf{x}}_{k|k} \qquad j \neq 0 \\ C_{ji}(\hat{\mathbf{x}}_{k|k}, \mathbf{x}_k) &= b_{ji} + c_{ji}\left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right]^T\left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right] \qquad i \neq 0, j \neq 0 \end{aligned} \quad (5.19)$$

Using the cost function (5.18), the integral in (5.17) can now be expressed as

$$I_j = \int_{R_j} P(\theta_0|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\theta_0)\, G(\mathbf{z}_k, \hat{\mathbf{x}}_{k|k}) d\mathbf{z}_k \tag{5.20}$$

where

$$\begin{aligned} G(\mathbf{z}_k, \hat{\mathbf{x}}_{k|k}) = & \sum_{i=0}^{M} P(\theta_i|\mathbf{Z}_{k-1}) \int [b_{ji} - b_{0i}]\, p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k \\ & + \sum_{i=0}^{M} P(\theta_i|\mathbf{Z}_{k-1}) \int [c_{ji} - c_{0i}]\, [\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}]^T [\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}] \\ & \times p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k \end{aligned} \tag{5.21}$$

The first term on the right hand side of the above equation is independent of $\hat{\mathbf{x}}_{k|k}$ and can be excluded from consideration in determining the optimum estimator. Since it is assumed that the cost of making a wrong decision is greater than the cost of making a correct decision ($c_{ji} \geq c_{ii}$), $G(\mathbf{z}_k, \hat{\mathbf{x}}_{k|k})$ is always positive or zero . So if the integrand is minimized then the integral will also be minimized and the optimum value of $\hat{\mathbf{x}}^*_{k|k}$ is found by differentiating the integrand with respect to the estimate and equating it to zero.

$$\left. \frac{\partial G(\mathbf{z}_k, \hat{\mathbf{x}}_{k|k})}{\partial \hat{\mathbf{x}}_{k|k}} \right|_{\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}^*_{k|k}} = 0 \tag{5.22}$$

Carrying out this minimization the optimum value of the state given hypothesis H_j is determined from

$$\sum_{i=0}^{M} [c_{ji} - c_{0i}]\, P(\theta_i|\mathbf{Z}_{k-1}) \int [\mathbf{x}_k - \hat{\mathbf{x}}^*_{k|k}]\, p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k = 0 \tag{5.23}$$

which gives

$$\hat{\mathbf{x}}^*_{k|k} = \frac{\sum_{i=0}^{M} [c_{ji} - c_{0i}]\, P(\theta_i|\mathbf{Z}_{k-1}) \int \mathbf{x}_k\, p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k}{\sum_{i=0}^{M} [c_{ji} - c_{0i}]\, P(\theta_i|\mathbf{Z}_{k-1}) \int p(\mathbf{z}_k|\mathbf{x}_k, \theta_i) p(\mathbf{x}_k|\mathbf{Z}_{k-1}, \theta_i) d\mathbf{x}_k}. \tag{5.24}$$

It is observed that no estimation is performed under hypothesis H_0. Thus, the

density function $p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_0)$ is equal to zero, and

$$[c_{ji}-c_{0i}]\,P(\theta_0|\mathbf{Z}_{k-1})\int \mathbf{x}_k\,p(\mathbf{z}_k|\mathbf{x}_k,\theta_0)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_0)d\mathbf{x}_k=0. \tag{5.25}$$

Using this result in (5.24) yields

$$\hat{\mathbf{x}}^*_{k|k}=\sum_{i=1}^{M}\Gamma_i(\mathbf{z}_k)\hat{\mathbf{x}}_{k|k,\theta_i} \tag{5.26}$$

where

$$\Gamma_i(\mathbf{z}_k)=\frac{[c_{ji}-c_{0i}]P(\theta_i|\mathbf{Z}_{k-1})\int p(\mathbf{z}_k|\mathbf{x}_k,\theta_i)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_i)d\mathbf{x}_k}{\sum_{m=0}^{M}[c_{jm}-c_{0m}]P(\theta_m)\int p(\mathbf{z}_k|\mathbf{x}_k,\theta_m)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_m)d\mathbf{x}_k} \tag{5.27}$$

and

$$\hat{\mathbf{x}}_{k|k,\theta_i}=\frac{\int \mathbf{x}_k\,p(\mathbf{z}_k|\mathbf{x}_k,\theta_i)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_i)d\mathbf{x}_k}{\int p(\mathbf{z}_k|\mathbf{x}_k,\theta_i)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_i)d\mathbf{x}_k}. \tag{5.28}$$

From Bayes rule

$$p(\mathbf{z}_k|\mathbf{x}_k,\theta_i)p(\mathbf{x}_k|\mathbf{Z}_{k-1},\theta_i)=p(\mathbf{x}_k|\mathbf{Z}_k,\theta_i)p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_i), \tag{5.29}$$

and $\hat{\mathbf{x}}_{k|k,\theta_i}$ becomes

$$\hat{\mathbf{x}}_{k|k,\theta_i}=\frac{\int \mathbf{x}_k p(\mathbf{x}_k|\mathbf{Z}_k,\theta_i)d\mathbf{x}_k}{\int p(\mathbf{x}_k|\mathbf{Z}_k,\theta_i)d\mathbf{x}_k}. \tag{5.30}$$

This is the mean of the a *posteriori* density function of $\mathbf{x}_k$ given the measurements $\mathbf{Z}_k$ and hypothesis H_i. If it is now assumed that

$$[c_{ji}-c_{0i}]=[c_{ki}-c_{0i}], \qquad 0\le i,j,k\le M \tag{5.31}$$

that is, the cost of all errors is the same, then (5.27) becomes

$$\begin{aligned}\Gamma_i(\mathbf{z}_k) &= \frac{P(\theta_i|\mathbf{Z}_{k-1}) \int p(\mathbf{x}_k|\mathbf{Z}_k,\theta_i)p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_i)d\mathbf{x}_k}{\sum_{m=0}^{M} P(\theta_m|\mathbf{Z}_{k-1}) \int p(\mathbf{x}_k|\mathbf{Z}_k,\theta_m)p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_m)d\mathbf{x}_k} \\ &= \frac{P(\theta_i|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_i)}{\sum_{m=0}^{M} P(\theta_m|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_m)}\end{aligned} \tag{5.32}$$

The denominator is the marginal density of $\mathbf{z}_k$ given the measurements $\mathbf{Z}_{k-1}$, and it can be shown ([77], p.85) that $\Gamma_i(\mathbf{z}_k) = P(\theta_i|\mathbf{Z}_k)$, the *a posteriori* probability of θ_i given the measurements $\mathbf{Z}_k$. This gives the recursion for the *a posteriori* probability of hypothesis H_i as

$$P(\theta_i|\mathbf{Z}_k) = \frac{P(\theta_i|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_i)}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})}. \tag{5.33}$$

Substituting the likelihood ratio

$$\Lambda_i(\mathbf{z}_k) = \frac{p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_i)}{p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_0)} \tag{5.34}$$

into (5.33), the *a posteriori* probability becomes

$$P(\theta_i|\mathbf{Z}_k) = \frac{P(\theta_i|\mathbf{Z}_{k-1})\, \Lambda_i(\mathbf{z}_k)}{\sum_{m=0}^{M} P(\theta_m|\mathbf{Z}_{k-1})\, \Lambda_m(\mathbf{z}_k)} \tag{5.35}$$

Under the conditions (5.31) the optimal estimate becomes

$$\hat{\mathbf{x}}^*_{k|k} = \sum_{i=1}^{M} P(\theta_i|\mathbf{Z}_k)\, \hat{\mathbf{x}}_{k|k,\theta_i}. \tag{5.36}$$

So the optimal estimate is the sum of all of the conditional means weighted by the *a posteriori* probability of each hypothesis.

Equations (5.35) and (5.36) are verified by Lainiotis [59] for Gaussian distributions. The implementation of the above procedure involves the operation of

several extended Kalman filters in parallel - one filter for each possible value of the parameter θ_i. In addition to estimating the state $\hat{\mathbf{x}}_{k|k,\theta_i}$ at each iteration of the Kalman filter, the *a posteriori* probability of θ_i, $P(\theta_i|\mathbf{Z}_k)$, is also propagated. The optimal mean square estimate $\hat{\mathbf{x}}_{k|k}$ of $\mathbf{x}_k$ is the integral over all values of θ_i of the estimates from each of the filters weighted by the *a posteriori* probability of each value of θ_i.

The error covariance $P_{k|k}$ is given by

$$P_{k|k} = E\left\{[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}][\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}]^T|\mathbf{Z}_k\right\} \tag{5.37}$$

which can be determined from the relation

$$P_{k|k} = \int E\left\{[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}][\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}]^T|\mathbf{Z}_k, \theta\right\} p(\theta|\mathbf{Z}_k)d\theta. \tag{5.38}$$

Substituting (5.7) gives

$$\begin{aligned} P_{k|k} &= \sum_{i=1}^{M} E\left\{[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}][\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}]^T|\mathbf{Z}_k, \theta_i\right\} P(\theta_i|\mathbf{Z}_k) \\ &= \sum_{i=1}^{M} \Big[E\left\{[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k,\theta_i}][\mathbf{x}_k - \hat{\mathbf{x}}_{k|k,\theta_i}]^T|\mathbf{Z}_k, \theta_i\right\} \\ &\qquad + [\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k,\theta_i}][\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k,\theta_i}]^T\Big] P(\theta_i|\mathbf{Z}_k) \\ &= \sum_{i=1}^{M} \left[P_{k|k,\theta_i} + \Delta P_{k|k,\theta_i}\right] P(\theta_i|\mathbf{Z}_k) \end{aligned} \tag{5.39}$$

where it is noted that $P_{k|k}$ is not defined under hypothesis θ_0. $P_{k|k,\theta_i}$ is the usual variance which is recursively computed by the Kalman filter under hypothesis H_i. $\Delta P_{k|k,\theta_i}$ represents the price of model uncertainty. It represents the performance degradation, or additional error, due to the fact that the model, characterized by θ_i, may not match the actual system that produced the measurements.

The estimator in (5.36) is the weighted sum of least squares estimators with the weights being the a *posteriori* probabilities of the various hypotheses. Because the cost of incorrect decisions is the same regardless of the decision, this estimator is independent of the decision.

5.2 JD/E for Systems in Gaussian Noise

For the measurement model (5.1), if the process and measurement noise are assumed to be Gaussian, the density function $p(\mathbf{z}_k|\theta_i, \mathbf{Z}_{k-1})$ used in (5.33) can be expressed as

$$p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_i) = |S_{k|k-1,\theta_i}|^{-1/2} \exp\left\{-1/2 \tilde{\mathbf{z}}_{k|k-1,\theta_i} S_{k|k-1,\theta_i} \tilde{\mathbf{z}}^T_{k|k-1,\theta_i}\right\} \tag{5.40}$$

where

$$\tilde{\mathbf{z}}_{k|k-1,\theta_i} = \mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1,\theta_i}, \theta_i) \tag{5.41}$$

and

$$S_{k|k-1,\theta_i} = H_k(\hat{\mathbf{x}}_{k|k-1,\theta_i}, \theta_i) P_{k|k-1,\theta_i} H_k(\hat{\mathbf{x}}_{k|k-1,\theta_i}, \theta_i)^T + R_{k,\theta_i} \tag{5.42}$$

with

$$H_k(\hat{\mathbf{x}}_{k|k-1,\theta_i}, \theta_i) = \left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k, \theta_i)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k|k-1,\theta_i}} . \tag{5.43}$$

The a *posteriori* probability (5.35) now becomes

$$P(\theta_i|\mathbf{Z}_k) = \frac{P(\theta_i|\mathbf{Z}_{k-1})|S_{k|k-1,\theta_i}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\tilde{\mathbf{z}}_{k|k-1,\theta_i} S^{-1}_{k|k-1,\theta_i} \tilde{\mathbf{z}}^T_{k|k-1,\theta_i}\right\}}{\sum_{m=0}^{M} P(\theta_m|\mathbf{Z}_{k-1})|S_{k|k-1,\theta_m}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\tilde{\mathbf{z}}_{k|k-1,\theta_m} S^{-1}_{k|k-1,\theta_m} \tilde{\mathbf{z}}^T_{k|k-1,\theta_m}\right\}} \tag{5.44}$$

5.3 JD/E in Non-Gaussian Noise

Two non-Gaussian distributions that are be of interest in subsequent chapters are the Weibull and lognormal. The high order filter, developed in Chapter 3, is used to perform parameter estimation in non-Gaussian noise. In the case of Weibull measurement noise, the a *posteriori* pdf, given the estimate $\hat{\mathbf{x}}_{k|k-1,\theta_i}$, can be updated as follows for a scalar measurement model z_k

$$p_w(z_k|\mathbf{Z}_{k-1},\theta_i) = a\frac{1}{\sigma_w}\left(\frac{\tilde{z}_{k|k-1,\theta_i}+\mu_w}{\sigma_w}\right)^{a-1}\exp\left\{-\left(\frac{\tilde{z}_{k|k-1,\theta_i}+\mu_w}{\sigma_w}\right)^{a}\right\} \quad (5.45)$$

The parameter a is a known constant that controls the skewness of the distribution. When $a = 2$ the Rayleigh distribution results. μ_w is the mean of the noncentral Weibull distribution. The n^{th} noncentral moment of the Weibull distribution is given by

$$E[(\tilde{z}_{k|k-1,\theta_i}+\mu_w)^n] = \Gamma\left(\frac{a+n}{a}\right)\sigma_w^n \quad (5.46)$$

where $\Gamma(.)$ is the Gamma function. Since $E[\tilde{z}_{k|k-1,\theta_i}] = 0$,

$$\mu_w = \Gamma\left(\frac{a+1}{a}\right)\sigma_w \quad (5.47)$$

The variance of $\tilde{z}_{k|k-1,\theta_i}$ is given by

$$\begin{aligned} s_{k|k-1,\theta_i} &= E[\tilde{z}^2_{k|k-1,\theta_i}] \\ &= \sigma_w^2\left[\Gamma\left(\frac{a+2}{a}\right) - \Gamma\left(\frac{a+1}{a}\right)^2\right] \end{aligned} \quad (5.48)$$

Given the the parameter a and the variance $s_{k|k-1,\theta_i}$, σ_w can be found from (5.48), and μ_w can subsequently be found from (5.47).

If a scalar error $\tilde{z}_{k|k-1,\theta_i}$ is centrally distributed according to the lognormal

distribution, its density function is updated by

$$p_l(z_k|Z_{k-1},\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_l(\tilde{z}_{k|k-1,\theta_i}+\mu_l)}\exp\left\{-\frac{1}{2\sigma_l^2}(\ln(\tilde{z}_{k|k-1,\theta_i}+\mu_l)-\gamma)^2\right\} \tag{5.49}$$

The mean and variance of the lognormal distribution are

$$\begin{aligned}\mu_l &= E[\tilde{z}_{k|k-1,\theta_i}+\mu_l] \\ &= \exp(\gamma+\sigma_l^2/2)\end{aligned} \tag{5.50}$$

and

$$\begin{aligned}s_{k|k-1,\theta_i} &= E[\tilde{z}^2_{k|k-1,\theta_i}] \\ &= \exp(2\gamma+2\sigma_l^2)-\exp(2\gamma+\sigma_l^2)\end{aligned} \tag{5.51}$$

Thus, given the variance $s_{k|k-1,\theta_i}$ and the parameter γ for the lognormal distribution, σ_l can be obtained from (5.51), and μ_l can be subsequently obtained from (5.50).

5.4 JD/E with Model Uncertainty

Consider the situation in which several different hypotheses are to be evaluated where the measurement model for each hypothesis may not be a function of all the elements of the state vector $\mathbf{x}_k$. For example, let $\mathbf{x}_k$ consist of two subvectors $\mathbf{x}_{1_k}$ and $\mathbf{x}_{2_k}$, such that $\mathbf{x}_k = [\mathbf{x}_{1_k}^T \vdots \mathbf{x}_{2_k}^T]^T$. Furthermore, let the measurement equation as a function of $\theta_i = [\vartheta_{i1}\,\vartheta_{i2}]^T$ be given by

$$\begin{aligned}H_i \quad : \quad \mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k,\theta_i)+\mathbf{v}_k \\ &= \vartheta i1\mathbf{h}_{1_k}(\mathbf{x}_{1_k})+\vartheta i2\mathbf{h}_{2_k}(\mathbf{x}_{2_k})+\mathbf{v}_k\end{aligned} \tag{5.52}$$

Thus $\mathbf{h}_k$ is segmented into two separate models, $\mathbf{h}_{1_k}$ and $\mathbf{h}_{1_k}$. Each model is a function of a subset of the state $\mathbf{x}_k$. Let us consider four hypotheses. Under hypothesis H_0 no signal is present. Under hypotheses H_1 and H_2, signals $\mathbf{h}_{1_k}$ and $\mathbf{h}_{2_k}$ are present, respectively. Hypothesis H_3 represents the situation where both

signals are present. The possible values of the parameter vector θ corresponding to the four different hypotheses are given by $\theta_0 = [00]^T$, $\theta_1 = [10]^T$, $\theta_2 = [01]^T$, and $\theta_0 = [11]^T$. The goal is to find the model that best fits the received data $\mathbf{z}_k$. For this situation it is not proper to use the optimal estimate represented by equation (5.36) since all state variables cannot be estimated by each model. Under these conditions the maximum a *posteriori* (MAP) decision rule is used. The model θ_i is chosen such that $P(\theta_i|\mathbf{Z}_k) = \max\{P(\theta_j;\ j = 0, \cdots, M\}$, where $M + 1$ is the number of hypotheses, including the null hypothesis. Note that for the example given above, if H_0 is chosen no estimate is required. If H_1 is chosen, then $\mathbf{x}_{1_k}$ can be estimated. If H_2 is chosen, then $\mathbf{x}_{2_k}$ can be estimated, and if H_3 is chosen,the full state vector $\mathbf{x}_k$ is estimated. Chapter 6 uses MAP estimation to perform model order selection for a system of sinusoids in Gaussian and non-Gaussian noise.

If the dimensionality of the state vector is different between models, or if state variable assignments are different between the various hypotheses, or if some of the state variables are unobservable between models, then the optimal estimate, represented by equation (5.36) no longer applies. That is, if not all variables can be estimated under each hypothesis (i.e. under each model), then it is not proper to form a combined estimate by summing the estimates from each model weighted by the a *posteriori* probability of each hypothesis.

5.5 JD/E with Uncertain Initial Conditions

Joint detection/estimation theory may also be applied to systems in which the measurement model is the same among all of the hypotheses but in which the different hypotheses are used to distinguish different sets of initial conditions. This can be very useful for nonlinear estimation problems since it is well known that

the performance of approximate estimators such as the extended Kalman filter is susceptable to errors in the initial estimates. This is due primarily to the first order Taylor series expansion used in the filter.

Consider the scalar case in which the initial estimation error on the state variable is uniformly distributed in $[-w/2,\ w/2]$ with mean $\hat{x}_{0|0}$. Let the width of the uniform distribution be w. The initial variance is $p_{0|0} = w^2/12$. Now consider the situation in which two models are used for the initial conditions, with each model accounting for one half of the uncertainty in the original model. In the first model, represented by θ_1, the initial estimate has mean $x_{0|0,\theta_1} = \hat{x}_{0|0} - w/4$. This mean is in the center of the left half of the original distribution for $\hat{x}_{0|0}$. The width of the uniform distribution for model 1 is $w/2$, and the initial estimation error variance is $p_{0|0,\theta_1} = w^2/48$. Similarly the mean and variance of the initial estimation error for model 2 are $x_{0|0,\theta_2} = \hat{x}_{0|0} + w/4$, and $p_{0|0,\theta_2} = w^2/48$.

For this example let $\theta_1 = -1$, and $\theta_2 = 1$. The two hypotheses can then be represented as

$$\begin{aligned} H_i \quad : \quad & z_k = h_k(x_k) + v_k \\ & x_{0|0,\theta_i} = \hat{x}_{0|0} + \theta_i w/4 \\ & p_{0|0,\theta_i} = w^2/48 \end{aligned} \tag{5.53}$$

With this model the performance of the extended Kalman filter is likely to be significantly more stable, as the initial estimation error variance is reduced by a factor of 4 compared to the original model. Since the state variables are the same for each model the optimal estimate described by (5.36) can be implemented. The a *priori* probabilities $P(\theta_i)$ are obtained by integrating the original initial density function over the limits used to partition the initial error into the separate hypotheses.

For the example given above $P(\theta_1) = P(\theta_2) = 0.5$.

This procedure could be particularly useful whenever the density function for the initial estimation error is multimodal. A filter could be constructed for each mode of the density function thereby greatly reducing the initial error variance.

As the number of partitions increases the joint detection/estimation problem becomes one of detection only. That is, the initial estimation error becomes so small that the implementation of the filter does not improve the estimate. Thus there is a tradeoff between estimation accuracy and the computational burden imposed by the implementation of several extended Kalman filters in parallel.

Chapter 7 uses joint detection/estimation for estimation of radar/sonar signal parameters in which the measurement model is the same for each hypothesis, but the initial estimates of the time delay and Doppler shift distinguish the hypotheses.

5.6 JD/E with Model Uncertainty and Uncertain Initial Conditions

The two estimation procedures discussed in the previous two sections can be used together to perform multiple hypothesis testing and for each hypothesis to have several sets of initial estimates. An application of this is given in Chapter 8 where radar/signal parameters are estimated from multiple sensor measurements.

5.7 Summary

A general procedure for joint detection/estimation has been presented. It is shown that this procedure may be used to segment the initial conditions of a estimation problem effectively controlling unstable behavior that characterizes nonlinear filtering techniques such as the extended Kalman filter in the presence of large initial

uncertainty. It is also shown that the joint detection/estimation procedure can be used for estimation problems with model uncertainty. In the following chapters this procedure is applied to specific signal processing problems.

Chapter 6

Joint Detection/Estimation for Model Order Selection

This chapter presents a general approach to determining the number of sinusoids present in measurements corrupted by additive white Gaussian and non-Gaussian noise. The approach involves the simultaneous application of maximum *a posteriori* (MAP) detection and nonlinear estimation using either the extended Kalman filter when the noise is Gaussian, or the extended high order filter (EHOF) when the noise is in non-Gaussian. The problem is formulated as a multiple hypothesis testing problem with assumed known *a priori* probabilities for each hypothesis. Each hypothesis represents a different measurement model. The unknown parameters for each model are estimated recursively along with the a posteriori probability of the hypothesis. The general technique for joint detection/estimation is presented in Chapter 5.

Other order selection methods [60 - 62] take the form of a function of the hypothesized number of parameters which penalizes over-estimation of the actual number of parameters when added to the log-likelihood function. A technique described Fuchs [63] uses eigenvector decomposition of the estimated autocorrelation matrix and is based on matrix perturbation analysis. In all of the autocorrelation techniques the additive noise is assumed to be Gaussian. Rao and Vaidyanathan [64] use cumulant based approach to estimate model order in non-Gaussian noise. In contrast to these methods, the technique used in this chapter is based on Bayes' theorem. The advantage of this technique is that it can be used in both Gaussian and non-Gaussian noise. It is completely general in that it applies to arbitrary *density functions.*

A typical method for determining the number of sinusoids present in a received signal is to form a model using all of the bins in the FFT as the maximum number of sinusoids present in the signal. If it is assumed that there is one sinusoid present in the measurement, then the number of hypotheses to be tested is N (the number of bins in the FFT). If there are two sinusoids present, then $N!/(2(N-2)!)$ hypotheses must be tested. If it is unknown whether one or two sinusoids are present, then $N + N!/(2(N-2)!)$ hypotheses must be tested. The obvious disadvantage of this approach is the exponential computational complexity in testing all hypotheses. In addition, the resolution of the frequencies is limited by the bin size of the FFT.

The method used in this chapter assumes that, if one sinusoid is present, then the estimates of the amplitude and frequency are known within some known mean and variance, that is, the distribution of the initial estimation error is assumed to be known. The procedure also allows for time varying variables (not allowed in the FFT method).

Simulation results are presented for the estimation of up to four sinusoids in white Gaussian, and non-Gaussian noise, when the actual number is two. In Gaussian noise the extended Kalman filter is used to perform estimation. In non-Gaussian noise the high order filter (EHOF) developed in Chapter 3 is used to perform estimation.

6.1 Joint Detection/Estimation Applied to Model Order Selection

The problem of model order selection can be cast into the framework of joint detection/estimation with model uncertainty. Section 5.4 describes the general solution for joint detection/estimation problems with model uncertainty. Consider the situation in which several different hypotheses are to be evaluated where

the measurement model for each hypothesis may not be a function of all of the state variables in the vector $\mathbf{x}_k$. For example, let $\mathbf{x}_k$ consist of P subvectors such that $\mathbf{x}_k = [\mathbf{x}_{1_k}^T : \mathbf{x}_{2_k}^T : \cdots : \mathbf{x}_{P_k}^T]^T$. Define the binary-element parameter vector $\theta_i \equiv [\vartheta_{i_1}\, \vartheta_{i_2} \cdots \vartheta_{i_P}]^T$, with $\vartheta_{i_p} = 0 \text{ or } 1$. The measurement model is given by

$$\begin{aligned} H_i \quad : \quad \mathbf{z}_k &= \mathbf{g}_k(\mathbf{x}_k, \theta_i) + \mathbf{v}_k \\ &= \sum_{p=1}^{P} \vartheta_{i_p} \mathbf{g}_{p_k}(\mathbf{x}_{p_k}) + \mathbf{v}_k \end{aligned} \tag{6.1}$$

with plant equation

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \theta) + \mathbf{w}_{k-1,\theta} \tag{6.2}$$

with initial state estimate $\hat{\mathbf{x}}_{0|0,\theta}$, and initial state covariance $P_{0|0,\theta}$. The initial state estimate, the measurement noise, and the process noise are uncorrelated. The process and measurement noise are zero mean and distributed with covariances $E[\mathbf{w}_{k,\theta}\mathbf{w}_{k,\theta}^T] = Q_{k,\theta}$, and $E[\mathbf{v}_{k,\theta}\mathbf{v}_{k,\theta}^T] = R_{k,\theta}$.

Hence, $\vartheta_{i_p} = 1$ indicates the presence of the p_{th} term in the i^{th} model; and $\vartheta_{i_p} = 0$ indicates its absence from the model. $\mathbf{g}_k$ is segmented into P separate models, with each model being a function of some subset of the state $\mathbf{x}_k$. Under hypothesis H_0 no signal is present and $\theta_0 = [0\,0 \cdots 0]^T$. There are P different possible combinations of one model only. The number of different combinations for more than one model is obtained from the binomial expansion. The total number of different models that can be accommodated by the measurement equation (6.1) is

$$N_p = 1 + \sum_{p=1}^{P} \frac{P!}{p!(P-p)!}$$

The goal is to find the model that best fits the received data $\mathbf{z}_k$, i.e. to select the parameter vector that gives the best fit. Given the measurements modeled as

(6.1), for model order selection it is not proper to use the MMSE estimate represented by equation (5.36), since all state variables cannot be estimated by each model. Under these conditions the maximum *a posteriori* decision rule should be used for model selection according to:

$$\text{Choose} \quad H_i \quad : \quad \theta_i = \text{argmax}_{\theta_m \in \Theta} P(\theta_m | \mathbf{Z}_k) \qquad m = 0, \cdots, M \quad (6.3)$$

where $M+1$ is the total number of hypotheses tested. The recursion for $P(\theta_m|\mathbf{Z}_k)$ from Chapter 5 is

$$P(\theta_i|\mathbf{Z}_k) = \frac{P(\theta_i|\mathbf{Z}_{k-1})\,\Lambda_i(\mathbf{z}_k)}{\sum_{m=0}^{M} P(\theta_m|\mathbf{Z}_{k-1})\,\Lambda_m(\mathbf{z}_k)} \quad (6.4)$$

where $\Lambda_i(\mathbf{z}_k)$ is the likelihood ratio

$$\Lambda_i(\mathbf{z}_k) = \frac{p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_i)}{p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_0)}, \quad (6.5)$$

and where $\mathbf{Z}_{k-1} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots \mathbf{z}_{k-1}\}$. The initial condition for (6.4) is the *a priori* probability density function $p(\theta) \equiv p(\theta|\mathbf{Z}_0)$, which is assumed to be known. The conditional probabilities $p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_i)$ are updated using the EKF or the EHOF as described below.

6.2 General System Model for Model Order Selection

Consider the problem of estimating the parameters of P unknown sinusoids from K measurements. The scalar measurement model for hypothesis H_i is given by

$$z_k = \sum_{p=1}^{P} \vartheta_{ip} c_{p_k} \exp(-\alpha_{p_k} k) \sin(\omega_{p_k} k + \phi_{p_k}) + v_k \quad (6.6)$$

for $k = 0, 1, \cdots, K-1$. It is assumed that the frequencies ω_{p_k} are normalized

so that the effective sampling interval is one second. v_k is assumed to be a white noise sequences each with variance σ^2. The objective is to estimate some or all of the $4P$ possibly time varying parameters in this system based on the measurements. Define the elements of the state variable subvector $\mathbf{x}_{p_k}$ as

$$\begin{aligned} x_{p_k}(1) &= \omega_{p_k} \\ x_{p_k}(2) &= c_{p_k} \\ x_{p_k}(3) &= \phi_{p_k} \\ x_{p_k}(4) &= \alpha_{p_k}. \end{aligned} \tag{6.7}$$

6.3 Model Order Selection Experimental Evaluation

In this section the performance of the joint detection/estimation method is evaluated experimentally. The number of sinusoids is unknown except for an upper bound. Furthermore, it is assumed that the damping coefficients and phases are all equal to zero. The amplitudes and frequencies are assumed to be either known or unknown. When they are unknown, estimates of them are obtained along with the model order selection. Assuming the unknown number of sinusoids to be four, the measurement equation becomes

$$\begin{aligned} \mathbf{z}_k &= \mathbf{g}_k(\mathbf{x}_k, \theta_i) + \mathbf{v}_k \\ &= \sum_{p=1}^{4} \vartheta_{ip} \mathbf{g}_{pk}(\mathbf{x}_{p_k}) + \mathbf{v}_k \end{aligned} \tag{6.8}$$

where

$$\mathbf{g}_{pk}(\mathbf{x}_{p_k}) = c_{p_k} \sin(\omega_{p_k} k) \tag{6.9}$$

The state variables are defined as

$$\begin{aligned} x_{p_k}(1) &= \omega_{p_k} \\ x_{p_k}(2) &= c_{p_k} \end{aligned} \tag{6.10}$$

The states are assumed to be constant with random initial values so that the plant equation becomes

$$\mathbf{x}_{k+1|k} = \mathbf{x}_k. \tag{6.11}$$

Since $P_{k|k-1,\theta_i} = P_{k-1|k-1,\theta_i}$ the extrapolation equation is redundant and the EKF equations (section 2.4.1) simplify to

$$\begin{aligned} \tilde{\mathbf{z}}_{k|k-1,\theta_i} &= H_k \tilde{\mathbf{x}}_{k|k-1,\theta_i} + \mathbf{v}_k \\ \hat{\mathbf{x}}_{k|k,\theta_i} &= \hat{\mathbf{x}}_{k-1|k-1,\theta_i} + K_k \tilde{\mathbf{z}}_k \\ S_{k|k-1,\theta_i} &= G_k P_{k-1|k-1,\theta_i} G_k^T + R_k \\ K_k &= P_{k-1|k-1,\theta_i} G_k^T S_{k|k-1,\theta_i}^{-1} \\ P_{k|k,\theta_i} &= (I_n - K_k G_k) P_{k-1|k-1,\theta_i} \end{aligned} \tag{6.12}$$

where the first partial derivative of the measurement nonlinearity (6.8, 6.9) under hypothesis H_i is given by

$$G_k = \left[\vartheta_{i1} \frac{\partial \mathbf{g}_{1_k}}{\partial \mathbf{x}_{1_k}} : \vartheta_{i2} \frac{\partial \mathbf{g}_{2_k}}{\partial \mathbf{x}_{2_k}} : \vartheta_{i3} \frac{\partial \mathbf{g}_{3_k}}{\partial \mathbf{x}_{3_k}} : \vartheta_{i4} \frac{\partial \mathbf{g}_{4_k}}{\partial \mathbf{x}_{4_k}} \right] \tag{6.13}$$

where

$$\frac{\partial \mathbf{g}_{pk}}{\partial \mathbf{x}_{p_k}} = \left[\begin{matrix} x_{p_k}(2)\, k \cos(x_{p_k}(1)\, k) \\ \sin(x_{p_k}(1)\, k) \end{matrix} \right]^T_{\mathbf{x}_{p_k} = \hat{\mathbf{x}}_{p_{k-1|k-1,\theta_i}}} \tag{6.14}$$

For estimation in Gaussian noise $\tilde{\mathbf{z}}_{k|k-1,\theta_i}$ and $S_{k|k-1,\theta_i}$ from (6.12) are used in (5.44) for computation of the a *posteriori* probability (6.4).

If the noise is non-Gaussian, the EHOF equations from Chapter 3 are used for

estimation. For the EHOF the innovations vector $\tilde{\mathbf{z}}_{k|k-1,\theta_i}$ and its variance $S_{k|k-1,\theta_i}$ are obtained in the same manner as they are in the EKF (6.12). However, in the EHOF the update equations for the filter variance $P_{k|k-1,\theta_i}$ (3.30 or 3.54) are much more complicated than they are for the EKF. In the experimental analysis that follows the measurement noise was distributed according to Weibull and lognormal distributions. The form of the density functions $p(\mathbf{z}_k|\mathbf{Z}_{k-1},\theta_i)$, which is required in (6.5), is given in section 5.3 for these distributions.

Five separate models are considered in the experimental evaluation. In the first model it is assumed that no signal is present, and $\theta_0 = [0\ 0\ 0\ 0]$, corresponding to the null hypothesis. The other four models correspond to the hypotheses that one, two, three or four different sinusoids are present in the measurements. The parameter vectors for these models are given by

$$\begin{aligned} \theta_1 &= [1\ 0\ 0\ 0] \\ \theta_2 &= [1\ 1\ 0\ 0] \\ \theta_3 &= [1\ 1\ 1\ 0] \\ \theta_4 &= [1\ 1\ 1\ 1] \end{aligned} \tag{6.15}$$

The *a priori* probability is chosen to be the same for each model, i.e. $P(\theta_i|0) = 1/5,\ i = 0, \cdots, 5$. The measurements (6.9) are modeled using four sinusoids with amplitudes $c_{p_k} = 1$, for $p = 1, \cdots 4$, and normalized frequencies $\omega_{1_k} = 0.12 * 2\pi$, $\omega_{2_k} = 0.22 * 2\pi$, $\omega_{3_k} = 0.32 * 2\pi$, and $\omega_{4_k} = 0.42 * 2\pi$. In the actual data only the first two sinusoids are present, namely the ones with frequencies ω_{1_k} and ω_{2_k}. The hypotheses are indexed according to the number of sinusoids assumed present in the data. The actual model corresponds to hypothesis H_2.

Three separate scenarios are evaluated. In the first scenario it is assumed that

there is no initial estimation error in the state variables, and therefore, estimation is not required. This corresponds to the detection-only case. In the second scenario only the frequencies are estimated. For this scenario equation (6.9) is modified accordingly to include only frequency variables. In the third scenario, both frequencies and amplitudes are estimated. Performance of the technique is evaluated via Monte Carlo simulations for v_k having Gaussian, Rayleigh, and lognormal measurement noise. The parameter $a = 2$ is used for the Rayleigh distribution (equation (5.45)), and $\gamma = 0$ is used for the lognormal distribution (equation (5.49)). In the scenarios in which the EHOF is employed, the 2^{nd}, 3^{rd}, and 4^{th} order statistics of the measurement noise are used in the filter implementation.

The detection results for scenario 1, the detection-only case, are presented in Tables 6.1, 6.2, and 6.3 for detection in Gaussian, Rayleigh, and lognormal noise, respectively. These tables contain the number of detection decisions for each model. The column labeled $P(\theta_2|\mathbf{Z}_k)$ gives the average a *posteriori* probability of the hypothesis H_2 for those simulation runs which chose H_2 as having the highest a *posteriori* probability. The results are shown as a function of signal to noise ratio (SNR) and the probability density function (pdf) type used in computing the likelihood ratio in equation (6.0). The SNR is defined as $10\log(c_{p_k}^2/(2\sigma_n^2))$, where σ_n^2 is the measurement noise variance. Since the amplitude is equal to one for all sinusoids, the SNR is $10\log(1/(2\sigma_n^2))$. In Table 6.1 only the measurement noise is Gaussian and only the Gaussian pdf is used to propagate the a *posteriori* probability. In Table 6.2 the noise is Rayleigh, and the a *posteriori* probability is computed using both the Rayleigh and Gaussian densities. Table 6.3 shows the results in lognormal noise. Tables 6.2 and 6.3 illustrate the importance of choosing the proper density function to make the detection decision.

Table 6.1. MAP Decisions as a Function of SNR

Gaussian Noise - Detection Only

SNR(dB)	pdf Type	H_0	H_1	H_2	H_3	H_4	$P(\theta_2\|\mathbf{Z}_k)$
-5	Gaussian	3	15	171	10	1	0.865
0	Gaussian	0	2	196	2	0	0.991
5	Gaussian	0	0	200	0	0	1.0

Table 6.2. MAP Decisions as a Function of SNR

Rayleigh Noise - Detection Only

SNR(dB)	Pdf Type	H_0	H_1	H_2	H_3	H_4	$P(\theta_2\|\mathbf{Z}_k)$
-5	Gaussian	2	16	169	12	1	0.872
	Rayleigh	3	8	180	8	1	0.925
0	Gaussian	0	3	197	0	0	0.991
	Rayleigh	0	0	199	1	0	0.999
5	Gaussian	0	0	200	0	0	1.0
	Rayleigh	0	0	200	0	0	1.0

Table 6.3. MAP Decisions as a Function of SNR

Lognormal Noise - Detection Only

SNR(dB)	Pdf Type	H_0	H_1	H_2	H_3	H_4	$P(\theta_2\|\mathbf{Z}_k)$
-5	Gaussian	2	15	174	8	1	0.885
	Lognormal	0	0	200	0	0	0.999
0	Gaussian	0	2	197	1	0	0.992
	Lognormal	0	0	200	0	0	1.0
5	Gaussian	0	0	200	0	0	1.0
	Lognormal	0	0	200	0	0	1.0

For scenario 2 it is assumed that the signal amplitudes a_{p_k} are known. The frequencies ω_{p_k} are estimated for each model. The standard deviation of the initial estimation error for the frequency in each model is $\sigma = 0.1$. Table 6.4 shows the number of times each hypothesis is chosen as a function of signal to noise ratio (dB) when the measurement noise is Gaussian and the measurements are processed with the EKF. The EHOF gives the same results as the EKF whenever the noise is Gaussian. Tables 6.5 and 6.6 show the results whenever the measurement noise is Rayleigh and lognormal, respectively. The development of the EKF is based on the fact that the filter error is a first order function of the innovations process. Thus, only first and second order statistics are necessary for EKF implementation. Therefore the EKF provides an optimal solution in Gaussian noise (providing the Taylor series approximation is valid). Although the EKF does not give optimal performance in non-Gaussian noise, it is evaluated in Tables 6.5 and 6.6 in order to compare its performance to the EHOF in non-Gaussian noise. The EKF is evaluated

in two configurations in these tables. The rows in these tables labeled 'EKF_g' denote the performance of the EKF in which density function in (6.5) is evaluated using a Gaussian density. 'EKF_r' and 'EKF_l' correspond to the performance of the EKF in which the Rayleigh and lognormal density functions are used for computation of the likelihood ratio. The EHOF employs the appropriate pdf associated with the measurement noise.

This data demonstrates that the nonlinear filtering techniques can give excellent performance for model order selection. Tables 6.5 and 6.6 demonstrate that the detection error probability for the EHOF is lower than that for the EKF in non-Gaussian noise, especially when the EKF is used in conjunction with the Gaussian density function. The EKF performs much better whenever the proper (Rayleigh of lognormal) density function is used. Furthermore, the EHOF decides with a higher confidence than the EKF, as demonstrated by the a *posteriori* probability $P(\theta_2|\mathbf{Z}_k)$. This difference occurs primarily at low values of the SNR.

The EHOF performs better relative to the EKF in lognormal noise than it does in Rayleigh. This is due to the fact that the lognormal noise has a higher degree of skewness than does the Rayleigh noise. That is, the EHOF has more of an advantage whenever the higher order statistics are large relative to what they would be in Gaussian noise.

Table 6.4. MAP Decisions as a Function of SNR

Gaussian Noise - Frequencies Estimated

SNR(dB)	Filter	H_0	H_1	H_2	H_3	H_4	$P(\theta_2 \mid \mathbf{Z}_k)$
-5	EKF	46	45	107	2	0	0.852
0	EKF	4	15	180	1	0	0.980
5	EKF	0	0	200	0	0	1.0

Table 6.5. MAP Decisions as a Function of SNR

Rayleigh Noise - Frequencies Estimated

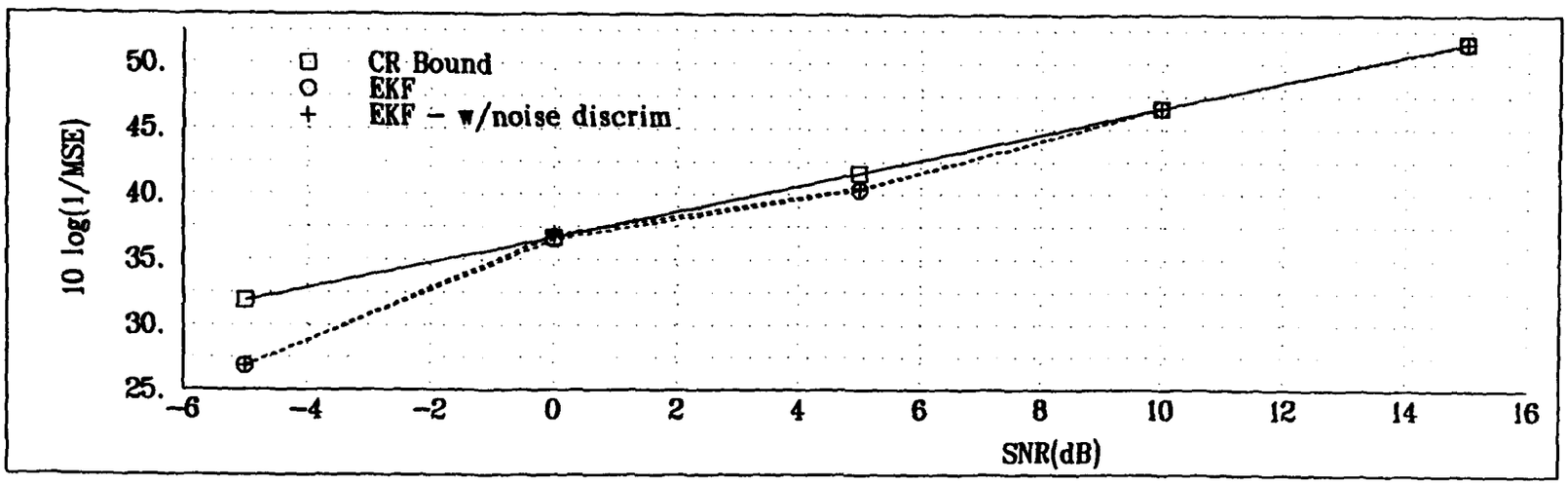
SNR(dB)	Filter	H_0	H_1	H_2	H_3	H_4	$P(\theta_2 \mid \mathbf{Z}_k)$
-5	EKF_g	44	46	108	2	0	0.844
	EKF_r	23	28	135	6	8	0.924
	EHOF	23	27	137	7	6	0.926
0	EKF_g	5	13	181	1	0	0.975
	EKF_r	2	16	174	8	0	0.996
	EHOF	1	11	182	6	0	0.995
5	EKF_g	0	0	200	0	0	1.0
	EKF_r	0	0	200	0	0	1.0
	EHOF	0	0	200	0	0	1.0

Table 6.6. MAP Decisions as a Function of SNR

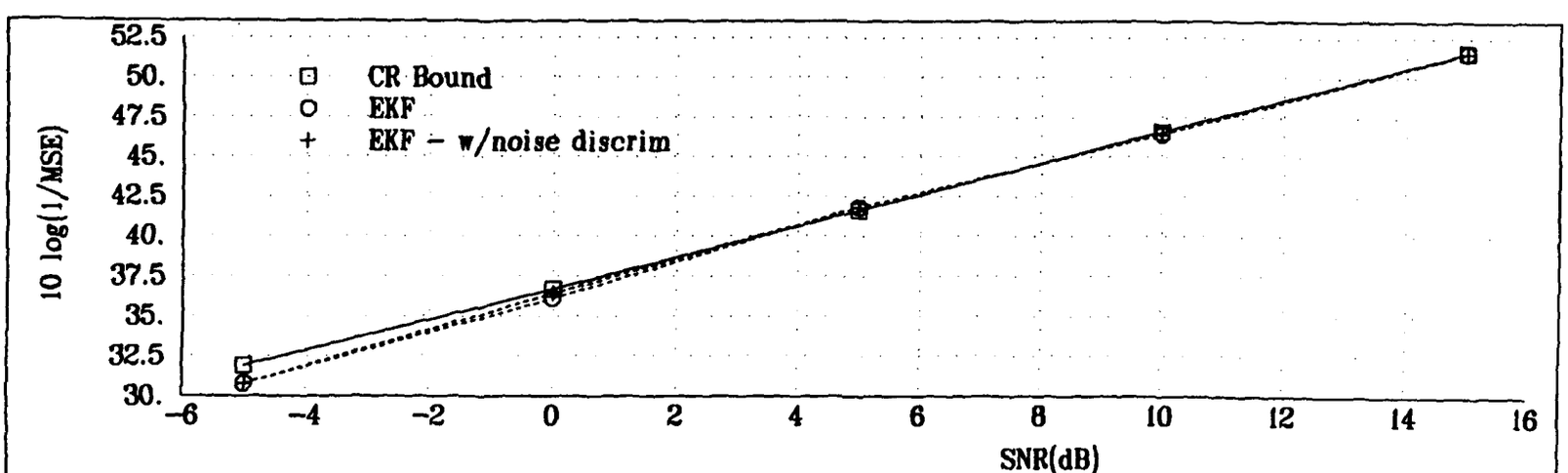
Lognormal Noise - Frequencies Estimated

SNR(dB)	Filter	H_0	H_1	H_2	H_3	H_4	$P(\theta_2\|\mathbf{Z}_k)$
-5	EKF_g	41	47	104	1	7	0.801
	EKF_l	27	30	129	11	3	0.934
	EHOF	24	20	146	5	5	0.954
0	EKF_g	3	13	183	1	0	0.979
	EKF_l	1	12	181	6	0	0.993
	EHOF	3	7	188	2	0	0.993
5	EKF_g	0	1	199	0	0	0.998
	EKF_l	0	1	198	1	0	0.999
	EHOF	0	0	200	0	0	1.0

Figures 6.1, 6.2, and 6.3 display the sample variance of the estimation error of the two estimated amplitudes as a function of SNR for estimation in Gaussian, Rayleigh, and lognormal noise. For Figure 6.1, the sample variance is computed only from those trials in the Monte Carlo simulation which resulted in the EKF choosing the correct hypothesis. Figures 6.2 and 6.3 display the sample variance for those trials which resulted in both the EKF and the EHOF choosing the correct hypothesis. The CR bound on the estimation error is also shown in these figures. A noise discrimination test is used in an attempt to detect poor estimates. This test involves discarding any estimate for which the sample variance of the residual $\tilde{z}_{k|k-1,\theta_i}$, computed over $k = 0, \cdots 24$, is greater than twice the noise variance σ_n^2. The results of using this test are also shown on Figures 6.1 - 6.3. The MSE of the

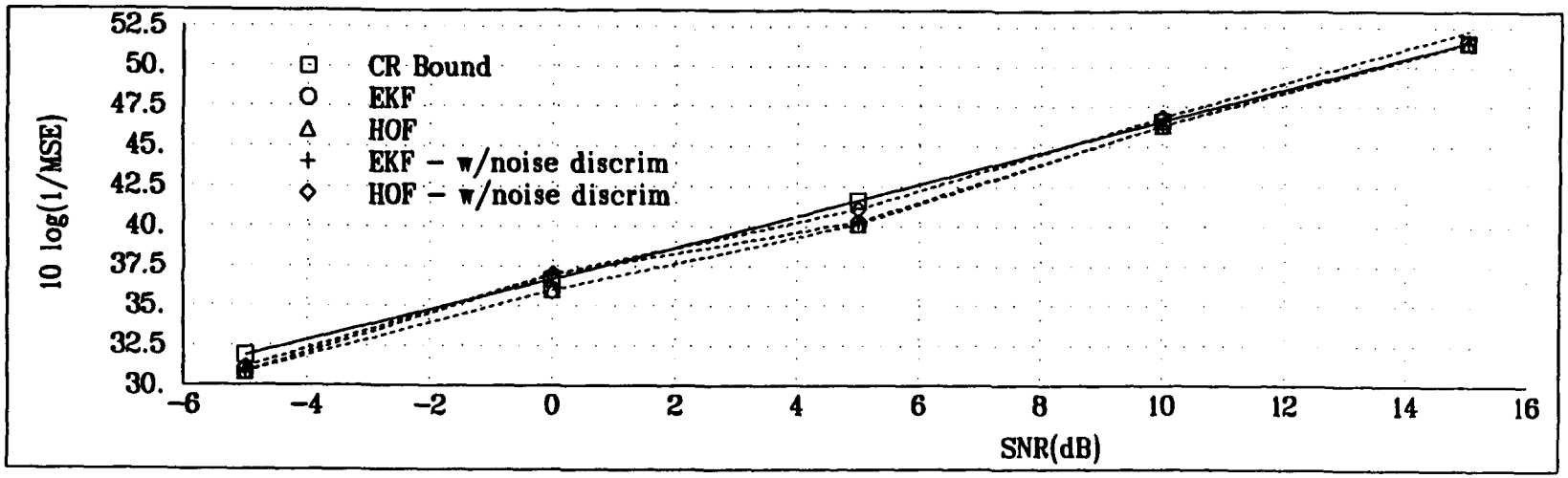


(a) Filter Performance for $\omega_{k_1} = 0.12 * 2\pi$

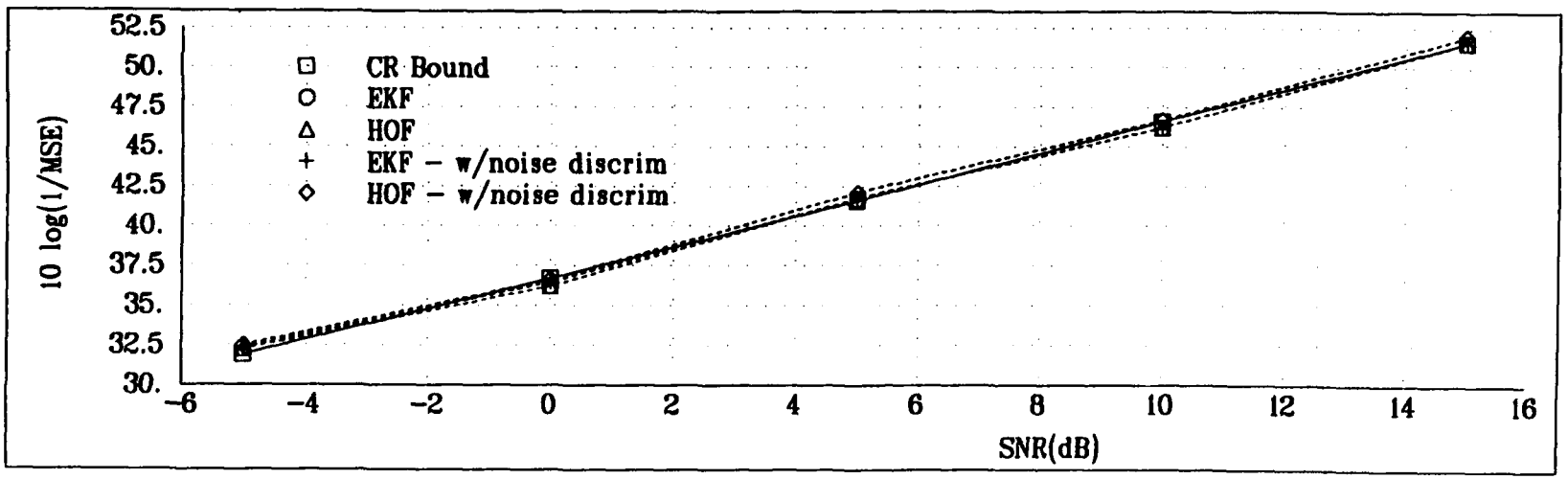


(b) Filter Performance for $\omega_{k_2} = 0.22 * 2\pi$

Figure 6.1 Nonlinear Filter Performance for the 2-State Model in Gaussian Noise, $P_0 = 0.01\ I$

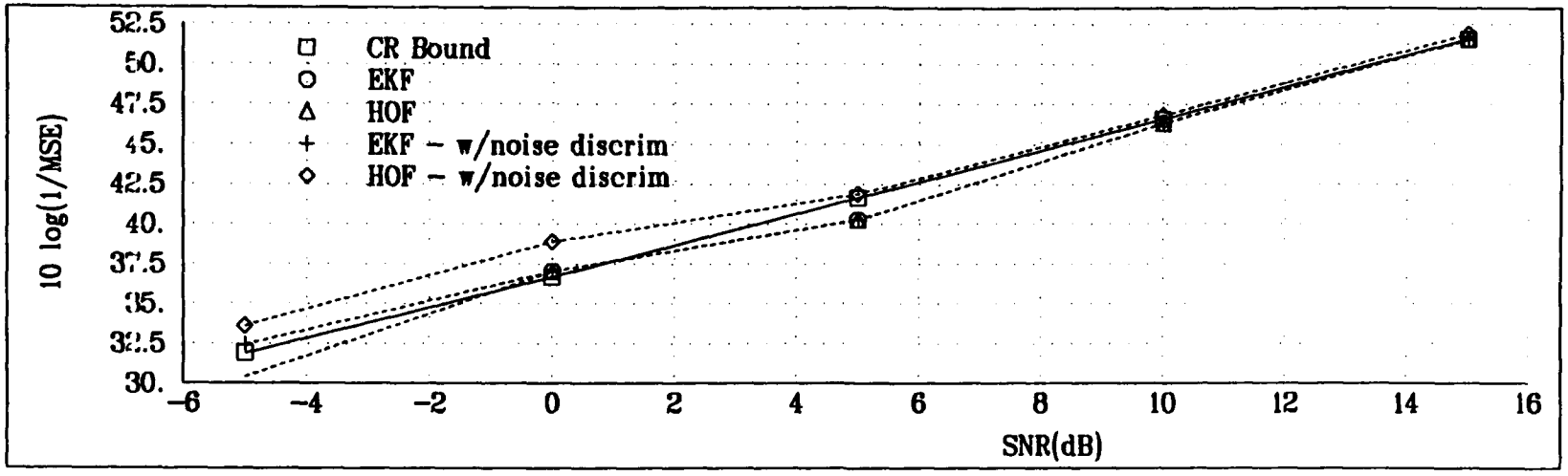


(a) Filter Performance for $\omega_{k_1} = 0.12 * 2\pi$

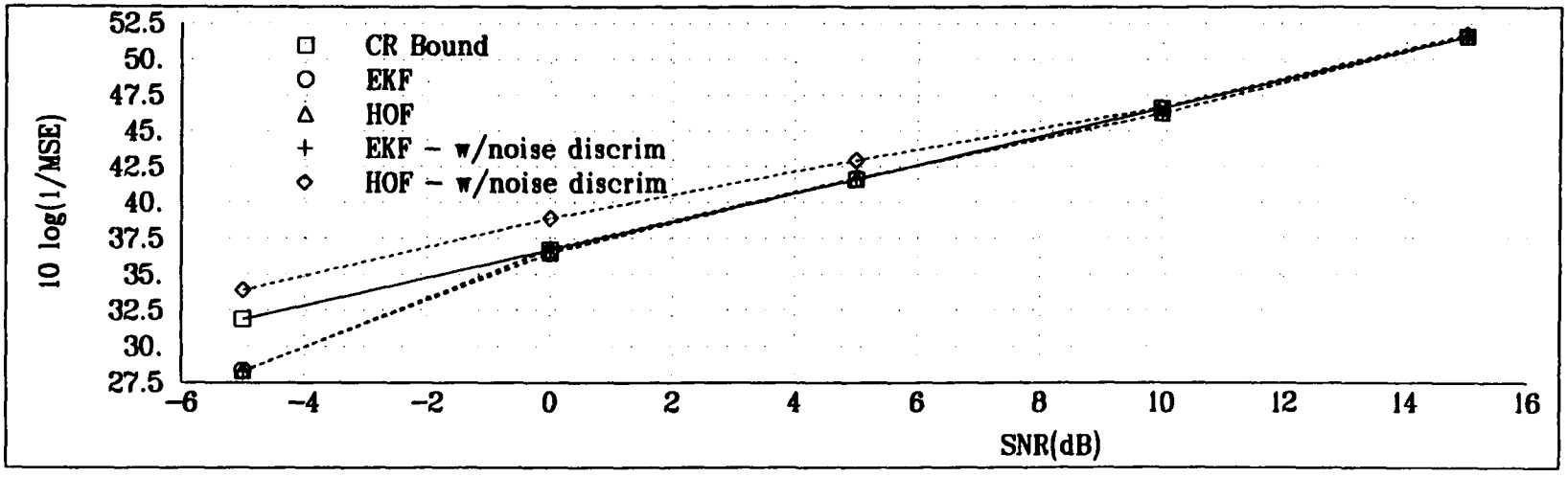


(b) Filter Performance for $\omega_{k_2} = 0.22 * 2\pi$

Figure 6.2 Nonlinear Filter Performance for the 2-State Model in Rayleigh Noise, $P_0 = 0.01\ I$



(a) Filter Performance for $\omega_{k_1} = 0.12 * 2\pi$



(b) Filter Performance for $\omega_{k_2} = 0.22 * 2\pi$

Figure 6.3 Nonlinear Filter Performance for the 2-State Model in Lognormal Noise, $P_0 = 0.01\ I$

EHOF is slightly lower than the EKF, with both filters producing results that are close to the Cramer Rao bound.

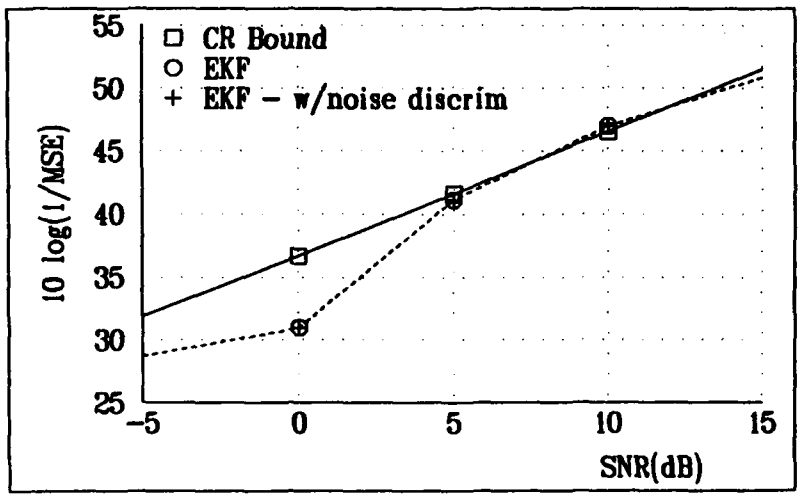
In scenario 3 both the signal amplitudes c_{p_k} and frequencies ω_{p_k} are estimated for each model. The standard deviation of the initial estimation is $\sigma = 0.1$ for both frequency and amplitude. Table 6.7 shows results for Gaussian noise, and Tables 6.8 and 6.9 display the results for Rayleigh and lognormal noise, respectively. These results are consistent with those in Tables 6.4 - 6.6 in that the EHOF makes better detection decisions than the EKF in non-Gaussian noise. However, as it can be expected, the probability of detection error increases whenever both frequency and amplitude are being estimated as compared to when only the frequency is estimated.

The estimation results for scenario 3 are given in Figures 6.4, 6.5, and 6.6 for estimation in Gaussian, Rayleigh, and lognormal measurement noises. Again it is shown that both the EKF and the EHOF perform close to the CR bound, with the EHOF giving better results than the EKF after the noise discrimination test is applied.

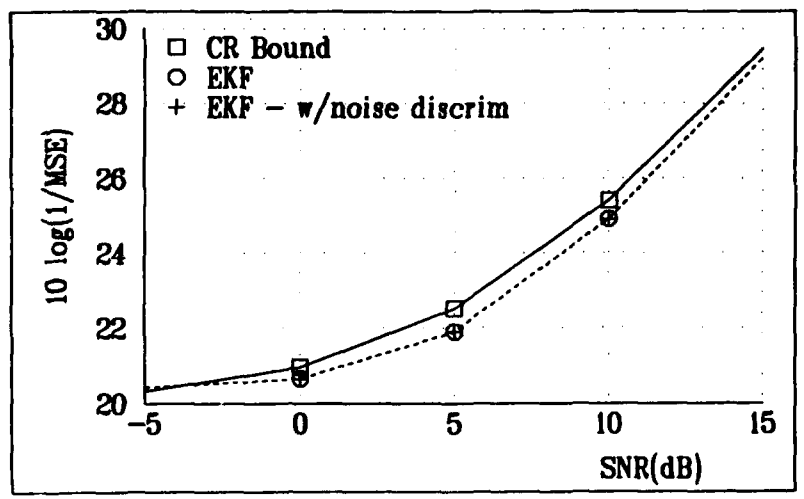
Table 6.7. MAP Decisions as a Function of SNR

Gaussian Noise - Amplitudes and Frequencies Estimated

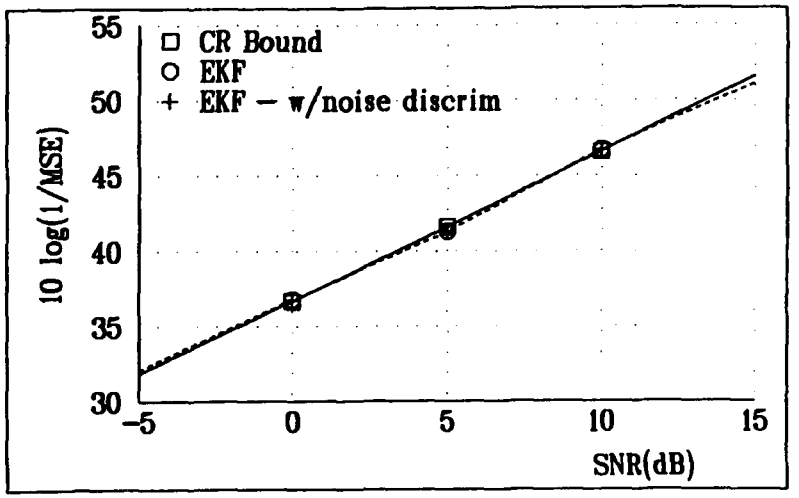
SNR(dB)	Filter	H_0	H_1	H_2	H_3	H_4	$P(\theta_2\|\mathbf{Z}_k)$
-5	EKF	49	51	95	5	0	0.792
0	EKF	1	25	174	0	0	0.971
5	EKF	0	0	200	0	0	1.0



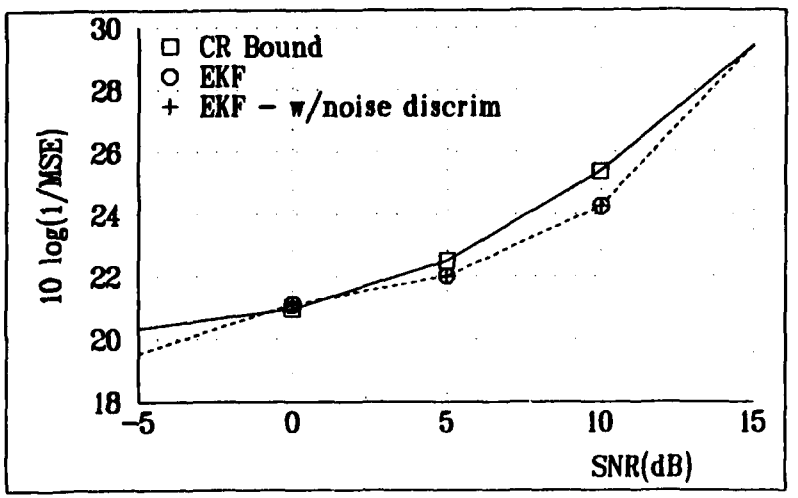
(a) Filter Performance for $\omega_{k_1} = 0.12 * 2\pi$



(b) Filter Performance for $c_{k_1} = 1.0$

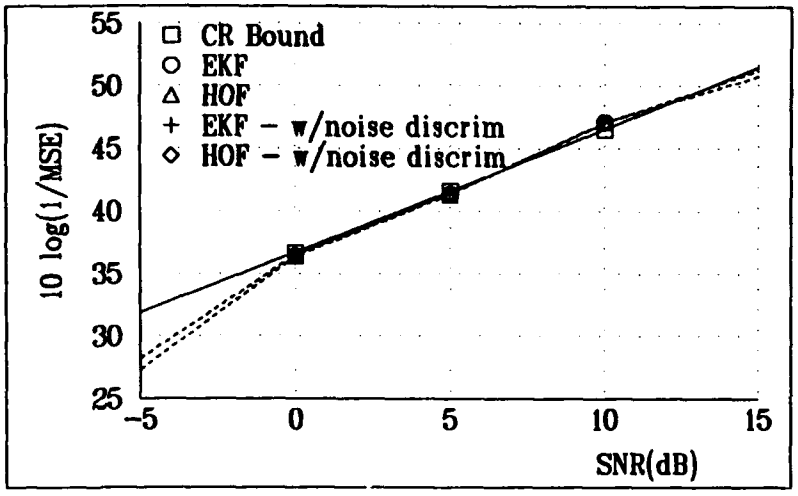


(c) Filter Performance for $\omega_{k_2} = 0.22 * 2\pi$

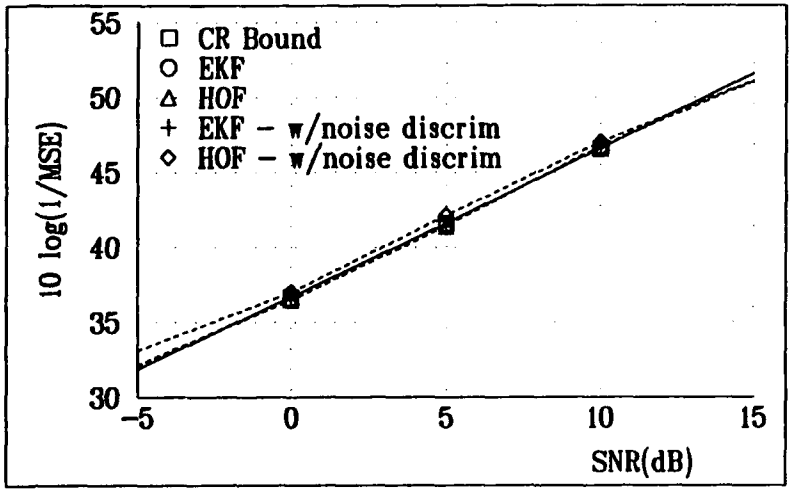


(d) Filter Performance for $c_{k_2} = 1.0$

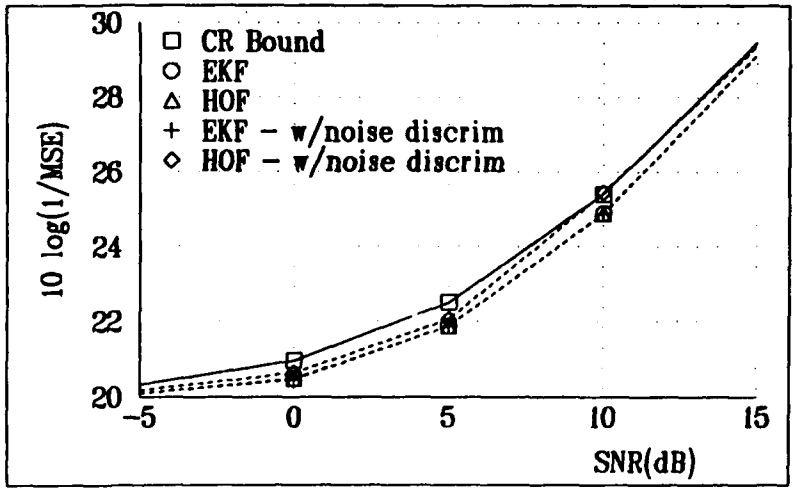
Figure 6.4 Nonlinear Filter Performance for the 4-State Model in Gaussian Noise, $P_0 = 0.01\ I$



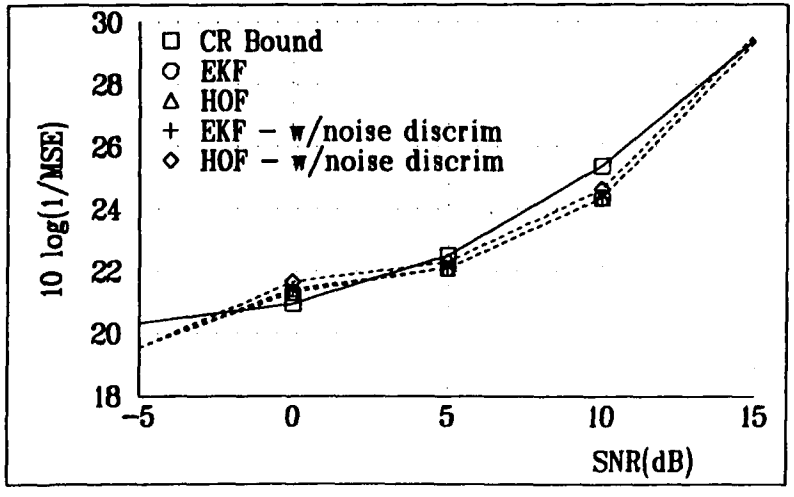
(a) Filter Performance for $\omega_{k_1} = 0.12 * 2\pi$



(c) Filter Performance for $\omega_{k_2} = 0.22 * 2\pi$

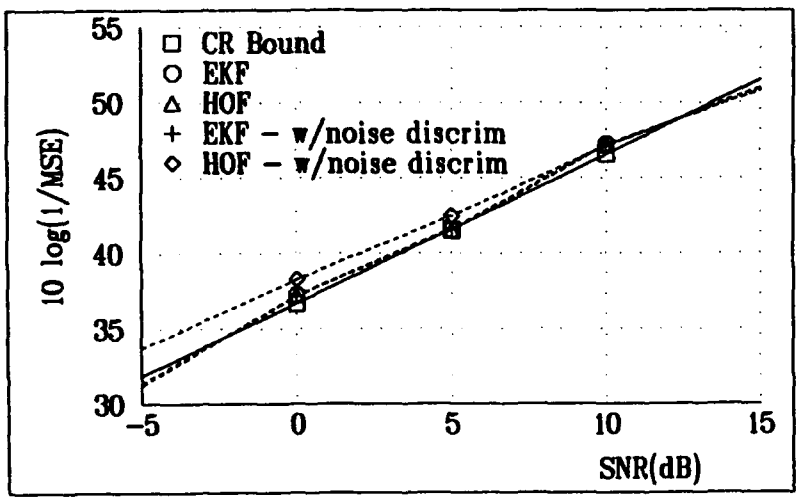


(b) Filter Performance for $c_{k_1} = 1.0$

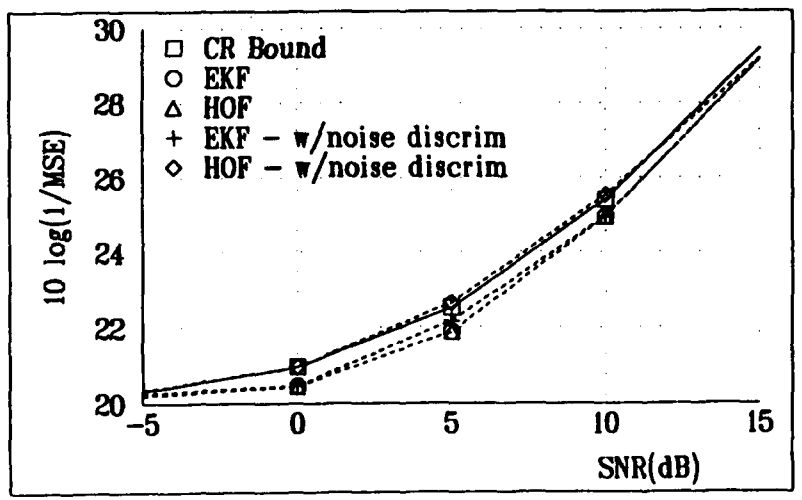


(d) Filter Performance for $c_{k_2} = 1.0$

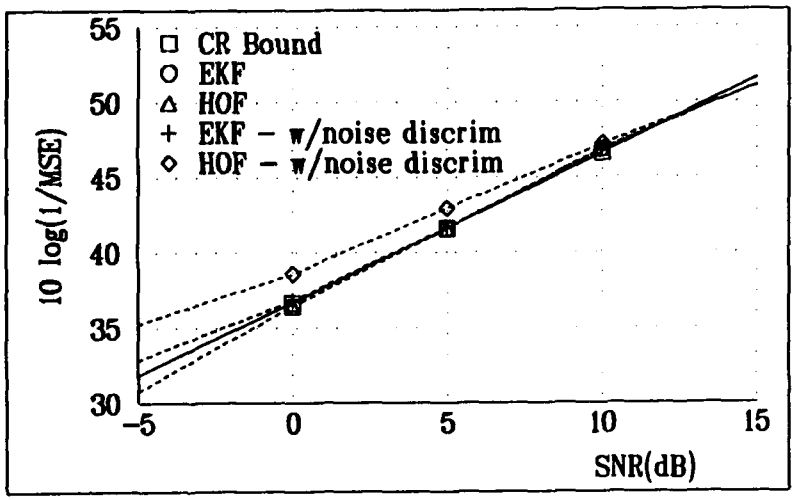
Figure 6.5 Nonlinear Filter Performance for the 4-State Model in Rayleigh Noise, $P_0 = 0.01\ I$



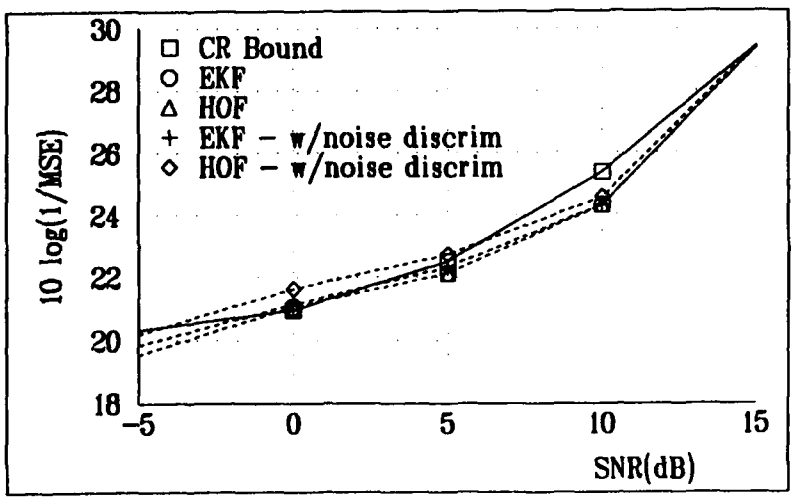
(a) Filter Performance for $\omega_{k_1} = 0.12 * 2\pi$



(b) Filter Performance for $c_{k_1} = 1.0$



(c) Filter Performance for $\omega_{k_2} = 0.22 * 2\pi$



(d) Filter Performance for $c_{k_2} = 1.0$

Figure 6.6 Nonlinear Filter Performance for the 4-State Model in Lognormal Noise, $P_0 = 0.01\ I$

Table 6.8. MAP Decisions as a Function of SNR

Rayleigh Noise - Amplitudes and Frequencies Estimated

SNR(dB)	Filter	H_0	H_1	H_2	H_3	H_4	$P(\theta_2\|\mathbf{Z}_k)$
-5	EKF_g	48	56	88	8	0	0.824
	EKF_r	20	31	125	19	5	0.880
	EHOF	20	34	124	20	2	0.885
0	EKF_g	1	25	174	0	0	0.973
	EKF_r	2	11	178	7	2	0.992
	EHOF	0	11	182	5	2	0.989
5	EKF_g	0	0	200	0	0	1.0
	EKF_r	0	0	200	0	0	1.0
	EHOF	0	0	200	0	0	1.0

Table 6.9. MAP Decisions as a Function of SNR

Lognormal Noise - Amplitudes and Frequencies Estimated

SNR(dB)	Filter	H_0	H_1	H_2	H_3	H_4	$P(\theta_2 \mid \mathbf{Z}_k)$
-5	EKF_g	44	52	98	6	0	0.783
	EKF_l	28	46	111	11	4	0.933
	EHOF	21	33	132	10	4	0.936
0	EKF_g	3	18	178	1	0	0.968
	EKF_l	1	14	179	6	0	0.995
	EHOF	0	7	184	8	1	0.995
5	EKF_g	0	0	200	0	0	1.0
	EKF_l	0	0	198	2	0	1.0
	EHOF	0	0	200	0	0	1.0

6.4 Conclusion

A general approach to model order selection has been presented based on joint detection/estimation theory. The approach involves the simultaneous application of maximum a *posteriori* detection theory and nonlinear estimation. The approach requires only an upper limit on the model order and is applicable to data that are being corrupted by additive Gaussian and non-Gaussian noise. The advantage of the approach lies in the potential to accommodate time varying as well as time invariant parameters in the measurement model. Experimental evaluation of the approach demonstrates excellent performance in selecting the correct model order and estimating the system parameters even in SNR's as low as -5 dB.

Chapter 7

JD/E Applied to Estimation of Time Delay and Doppler Shift

A nonlinear adaptive detector/estimator (NADE) is introduced for single and multiple sensor data processing. The problem of target detection from returns of monostatic sensor(s) is formulated as a nonlinear joint detection/estimation problem on the unknown parameters in the signal return. The unknown parameters involve the presence of the target, its range, azimuth, and Doppler velocity. The problems of detecting the target and estimating its parameters are considered jointly. A bank of spatially and temporally localized nonlinear filters is used to estimate the *a posteriori* likelihood of the existence of the target in a given space-time resolution cell. Within a given cell, the localized filters are used to produce refined spatial estimates of the target parameters. A decision logic is used to decide on the existence of a target within any given resolution cell based on the *a posteriori* estimates reduced from the likelihood functions. The inherent spatial and temporal referencing in this approach is used for automatic referencing required when multiple sensor data is fused together. Thus, the approach is naturally extendable to centralized multisensor data fusion.

This chapter addresses the joint estimation of time delay and Doppler shift from measurements of a received signal. Knapp and Carter [66] showed that the ML estimator of time delay can be represented by a pair of prefilters followed by a matched filter. Stuller [67] generalized these results to obtain ML estimates of time varying delay, nonstationary signals, and arbitrary observation interval. An extension of the ML methods is given by Abatzoglou [68] in which local maximization

of the cross-correlation function results in fast ML estimation. An optimum time delay tracker based on a first order markov model is given in [69]. In [70] the same model is used for optimum time delay detection and tracking. These studies assume linear plant and measurement equations so that an optimal solution can be obtained using the Bayesian approach.

A general overview of techniques employed for time delay estimation in sonar signal processing is given by Carter [71]. Stein [72] describes how the complex ambiguity function can be used for joint estimation of time delay and Doppler shift. The ambiguity function approach is one of the most widely accepted methods for joint estimation of time delay and Doppler shift. The major disadvantage of this approach is that its implementation requires the use of the Fourier transform. Thus the resolution is limited, especially for short data lengths. Time delay estimation has also been approached using higher order statistics. Nikias and Pan [73] and Chiang and Nikias [74] make use of the fact that Gaussian noise is suppressed in the third order cumulant domain to form estimates of time varying delay.

This chapter considers the problem of localizing a target in a range-Doppler space. The range-Doppler space is partitioned into a number of resolution cells. Each cell is identified with a hypothesis that the signal is present in it. A joint detection/estimation scheme is then used to localize the target and refine its parameter estimates (i.e. time delay and Doppler shift). The measurements that are used to localize the target consist of signal returns corrupted by additive white Gaussian and non-Gaussian noise.

The problem is formulated using the joint detection/estimation procedure developed in Chapter 5 adapted to problems with uncertain initial conditions. The

approach involves the operation of several nonlinear independent filters in parallel. In the case of Gaussian measurement noise the extended Kalman filter is used for estimation. The extended high order filter (EHOF), developed in Chapter 3, is used in non-Gaussian noise. The parallel filters are distinguished by the initial conditions used to set up the problem. Along with the state estimate the *a posteriori* probability of each hypothesis is computed recursively.

Two different implementations are evaluated. In the first implementation, the model parameters for each resolution cell are kept fixed at their *a priori* estimates. The fixed estimates are then used to update the *a posteriori* probability of each cell. In the second implementation, the model parameters for each resolution cell are estimated on-line and used to update the a *posteriori* probability for each resolution cell. After all data is processed, the *a posteriori* probabilities and the initial estimates are used to produce a minimum mean square error (MMSE) estimate of the time delay and Doppler shift.

7.1 Problem Statement

Consider the problem of signal detection and parameter estimation in the context of the reception of an active echo return from a object that has been illuminated by a monostatic source. The situation is considered in which there are P collocated sources that illuminate the target simultaneously, but with different carrier frequencies designated ω_{c_p}. The received signal at each sensor is frequency-translated by mixing it with a signal at frequency ω_{t_p}. The resulting signal is low-pass filtered, and digitized at a rate f_s, which is at least twice the highest frequency in the data. The time between samples is denoted t_s. It is assumed that all sensors have the same digitization rate, and that all clocks are synchronized. The

general expression for the received signal at the p^{th} sensor can be written

$$z_{kp} = a_{kp}(\tau_k)p_{kp}(\tau_k, \nu_k)r_{kp}(\tau_k, \nu_k) + v_{kp} \tag{7.1}$$

where $a_{kp}(\tau_k)$ is the received signal amplitude, $p_{kp}(\tau_k, \nu_k)$ is the pulse shaping function, and

$$r_{kp}(\tau_k, \nu_k) = \cos\left[(\nu_k(\omega_{cp}\,(k\,t_s - \tau_k))) - \omega_{tp}k\,t_s\right] \tag{7.2}$$

v_{kp} is white noise with $E[v_{kp}] = 0$, $E[v_{kp}v_{jp}] = \sigma_{kp}^2\delta(k-j)$. and τ_k is the time delay between signal transmission and reception. τ_k is a function of the range D_k between the receiver and the object, and is given by

$$\tau_k = \frac{2D_k}{c} \tag{7.3}$$

The Doppler shift parameter ν_k is given [76] by

$$\nu_k = 1 + \frac{2V_{d_k}}{c} \tag{7.4}$$

where V_{d_k} is the Doppler velocity obtained by projecting the velocity vectors of the target and receiver along the line of sight between them, and c is the speed of electromagnetic propagation. ν_k is bounded by the perceived maximum speed of the object and exact knowledge of the receiver platform speed. Based on these capabilities one could postulate fairly accurate representations for the moments of the probability density functions for ν_k. For unambiguous range estimation the uncertainty in τ_k, denoted $\Delta\tau_k$ is bounded by $\Delta\tau_k \leq 2\pi/(\nu_k\omega_{cp})$. This is due to the fact that the cos(.) function is not monotonic (i.e. $r_{kp}(\tau_1, \nu_k) = r_{kp}(\tau_2, \nu_k)$, if $\tau_2 - \tau_1 = 2\pi/(\nu_k\omega_{cp})$).

$p_{kp}(\tau_k, \nu_k)$ is the pulse shaping function, which has average energy E_p. The

signal amplitude is attenuated due to spherical spreading loss by a factor of $1/D_k^2$. With known transmitted amplitude A_{t_p}, the received amplitude as a function of time delay is given by

$$a_{k_p}(\tau_k) = \frac{4A_{t_p}}{(c\,\tau_k)^2} \tag{7.5}$$

Note that the effects of filtering, amplification, and digitization on the received signal amplitude are not considered. These effects are generally known and can be accounted for in the amplitude function (7.5).

7.2 Joint Detection/Estimation

The joint detection/estimation procedure for problems with uncertain initial conditions is followed in this chapter for optimal estimation of time delay and Doppler shift. This procedure is described in Section 5.5. The hypotheses are distinguished from each other by the initial conditions on the initial state estimates , $\hat{\mathbf{x}}_{0|0,\theta_i}$, and initial state covariances $P_{0|0,\theta_i}$. The measurement and process models are the same for each hypothesis. Let $\theta_i \in \Theta$ designate the parameter vector that describes the different initial conditions on the states. The parameter vector θ_i is also assumed to be time invariant. Under hypothesis H_{θ_i} the discrete time measurements are modeled according to

$$\begin{aligned} H_{\theta_i} \quad : \quad & \mathbf{z}_k = \mathbf{g}_k(\mathbf{x}_k) + \mathbf{v}_k \\ & \text{with i.c.'s } \hat{\mathbf{x}}_{0|0,\theta_i}, P_{0|0,\theta_i} \end{aligned} \tag{7.6}$$

The measurement vector $\mathbf{z}_k$ is composed of the scalar measurements of the P individual sensors such that

$$\mathbf{z}_k = \begin{bmatrix} z_{k_1} & z_{k_2} & \cdots & z_{k_P} \end{bmatrix}^T \tag{7.7}$$

The state $\mathbf{x}_k$ is common for all $\theta_i \in \Theta$, and satisfies the discrete time process equation

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1} \tag{7.8}$$

The initial state estimate, the measurement noise, and the process noise are uncorrelated. The process and measurement noise are zero mean and distributed with covariances $E[\mathbf{w}_k\mathbf{w}_k^T] = Q_k$, and $E[\mathbf{v}_k\mathbf{v}_k^T] = R_k$.

For each $\theta_i \in \Theta$ (each assumed model), a minimum variance estimate of the model parameters is obtained recursively using the joint detection/estimation technique. Using this technique a minimum variance estimate of the model parameters is obtained for every assumed model. These estimates are subsequently used to estimate the likelihood of each model being the correct one. Based on these likelihood estimates, a maximum *a posteriori* (MAP) decision criteria or a minimum mean square error (MMSE) decision criteria can be used to select the proper model.

Using Bayes' rule, the *a posteriori* probability of the parameter vector θ is updated recursively by [67, 68]

$$P(\theta_i|\mathbf{Z}_k) = \frac{P(\theta_i|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_i)}{\sum_{m=1}^{M} P(\theta_m|\mathbf{Z}_{k-1})\, p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_m)} \tag{7.9}$$

where $\mathbf{Z}_{k-1} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots \mathbf{z}_{k-1}\}$. The initial condition for (7.9) is the *a priori* probability density function $p(\theta) \equiv p(\theta|\mathbf{Z}_0)$, which is assumed to be known. The densities $p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_i)$ are updated using the EKF or the EHOF.

The update procedure for measurements in Gaussian, Rayleigh, and lognormal noise is described in Chapter 5, sections 5.2 and 5.3 .

Since the state vector $\mathbf{x}_k$ is common to all models, the minimum mean

squared error (MMSE) estimate can be used. The MMSE estimate is expressed as a weighted average of the conditional state estmates $\hat{\mathbf{x}}_{k|k,\theta_i}$ over all θ_i as follows:

$$\hat{\mathbf{x}}^*_{k|k} = \sum_{i=1}^{M} P(\theta_i|\mathbf{Z}_k)\,\hat{\mathbf{x}}_{k|k,\theta_i}. \tag{7.10}$$

7.3 Specification of Initial Conditions

The localized initial conditions for each resolution cell are defined as follows: Let the time delay have mean $\hat{\tau}_0$ and density function $p_{\tau_0}(\tau_0)$. The distribution of τ_0 is segmented into N nonoverlapping segments such that the segment around some localized initial estimate $\hat{\tau}_{n_0}$ is defined by

$$p_{\tau_{n_0}}(\tau_{n_0}) = p_{\tau_0}(\tau_0) \qquad \alpha_n \le \tau_0 \le \alpha_{n+1} \qquad 1 \le n \le N \tag{7.11}$$

We have

$$\sum_{n=1}^{N} \int_{\alpha_n}^{\alpha_{n+1}} p_{\tau_{n_0}}(\tau)d\tau = \int_{-\infty}^{\infty} p_{\tau_0}(\tau)d\tau = 1$$

Define the scaling parameters ζ_n such that

$$\zeta_n \int_{\alpha_n}^{\alpha_{n+1}} p_{\tau_{n_0}}(\tau)d\tau = 1 \qquad 1 \le n \le N$$

Then the mean and variance of the initial conditions of the segmented model are given by

$$\hat{\tau_{n_0}} = E[\tau_{n_0}] = \zeta_n \int_{\alpha_n}^{\alpha_{n+1}} \tau p_{\tau_{n_0}}(\tau)d\tau$$

$$\mathrm{Var}[\tau_{n_0}] = \zeta_n \int_{\alpha_n}^{\alpha_{n+1}} \tau^2 p_{\tau_{n_0}}(\tau)d\tau - \hat{\tau_{n_0}}^2$$

Similarly, the initial estimate of Doppler shift have mean $\hat{\nu}_0$ and density function $p_{\nu_0}(\nu_0)$. Now let the distribution for ν_0 be segmented into M nonoverlapping seg-

ments such that the segment corresponding to initial estimate ν_{m_0} is defined by

$$p_{\nu m_0}(\nu_{m_0}) = p_{\nu_0}(\nu_0) \qquad \gamma_m \leq \nu_0 \leq \gamma_{m+1} \qquad 1 \leq m \leq M \tag{7.12}$$

We have

$$\sum_{m=1}^{M} \int_{\gamma_m}^{\gamma_{m+1}} p_{\nu m_0}(\nu) d\nu = \int_{-\infty}^{\infty} p_{\nu_0}(\nu) d\nu = 1$$

Define the scaling parameters κ_m such that

$$\kappa_m \int_{\gamma_m}^{\gamma_{m+1}} p_{\nu m_0}(\nu) d\nu = 1 \qquad 1 \leq m \leq M$$

Then the mean and variance of the initial conditions of the segmented model for ν_0 are given by

$$\hat{\nu}_{m_0} = E[\nu_{m_0}] = \kappa_m \int_{\gamma_m}^{\gamma_{m+1}} \nu p_{\nu m_0}(\nu) d\nu$$

$$\text{Var}[\nu_{m_0}] = \kappa_m \int_{\gamma_m}^{\gamma_{m+1}} \nu_{m_0}^2 p_{\nu m_0}(\nu) d\nu - \hat{\nu}_{m_0}^2$$

With N different initial conditions on τ_0, and M different initial conditions on ν_0 there are NM different resolution cells for referencing the measurements. A different filter is initialized in each resolution cell. The total number of cells, MN, in the resolution space can be large, depending on the desired accuracy in the parameter resolution. However, the filters can be run in parallel, and independent of each other, thus reducing the execution time to that of a single filter.

The parameter vector θ_i, $i = (n-1) * M + m$, $1 \leq n \leq N$, $1 \leq m \leq M$, is defined to be the $(n,m)^{th}$ resolution cell and is used to define NM initial conditions on the state variables τ and ν. The *a priori* probabilities of each hypothesis are determined by integrating the density functions $p_{\tau_0}(\tau_0)$ and $p_{\nu_0}(\nu_0)$ over the limits

defined for each hypothesis. They are given by

$$P(\theta_i) = \int_{\alpha_n}^{\alpha_{n+1}} p_{\tau_0}(\tau)d\tau \int_{\gamma_m}^{\gamma_{m+1}} p_{\nu_0}(\nu)d\nu \tag{7.13}$$

7.4 Joint Detection/Estimation of Delay and Doppler

The model is now considered in which both ν_k and τ_k are unknown, with both state variables to be estimated. The parameter vector $\theta_i = [n\ m]^T, 1 \leq i \leq NM$ is used to define NM different initial conditions on the state variables τ and ν. The hypothesis H_i that corresponds to θ_i for sensor p is given by

$$H_i \quad : \quad z_{kp} = \begin{cases} v_k & k\,t_s < \hat{\tau}_k \\ g_{kp}(\hat{\tau}_k, \hat{\nu}_k) + v_k & \hat{\tau}_k \leq k\,t_s < \hat{\tau}_k + t_w \\ v_k & k\,t_s \geq \hat{\tau}_k + t_w \end{cases} \tag{7.14}$$

where t_w is the pulse width. The initial conditions are given by

$$\begin{aligned} \hat{x}_{0|0,\theta_i} &= [\hat{\tau}_{n_0}, \hat{\nu}_{m_0}]^T \\ P_{0|0,\theta_i} &= \begin{bmatrix} \mathrm{Var}[\tau_{n_0}] & 0 \\ 0 & \mathrm{Var}[\nu_{m_0}] \end{bmatrix} \end{aligned} \tag{7.15}$$

where

$$g_{kp}(\hat{\tau}_k, \hat{\nu}_k) = a_{kp}(\hat{\tau}_k) p_{kp}(\hat{\tau}_k, \hat{\nu}_k) r_{kp}(\hat{\tau}_k, \hat{\nu}_k) \tag{7.16}$$

$$
\begin{aligned}
a_{kp}(\hat{\tau}_k) &= \frac{4A_{t_p}}{(c\,\hat{\tau}_k)^2} \\
p_{kp}(\hat{\tau}_k, \hat{\nu}_k) &= 0.5\,(1 - \cos(2\pi\,\hat{\nu}_k(k\,t_s - \hat{\tau}_k)/t_w)) \\
r_{kp}(\hat{\tau}_k, \hat{\nu}_k) &= \cos\left[(\hat{\nu}_k(\omega_{cp}\,(k\,t_s - \hat{\tau}_k))) - \omega_{t_p} k\,t_s\right]
\end{aligned}
\tag{7.17}
$$

The Hanning window is used as the pulse shaping function $p_{kp}()$.

With the state variable vector as $\mathbf{x}_k = [\tau_k\ \nu_k]^T$, the Jacobian of the measurement model is given by

$$
\mathbf{G}_k = \left[\mathbf{G}_{k1}^T\ \mathbf{G}_{k2}^T\ \cdots\ \mathbf{G}_{kP}^T\right]^T \tag{7.18}
$$

where the Jacobian $\mathbf{G}_{kp}$ for the p^{th} sensor , that is used in both the EKF and the EHOF, is given by

$$
\begin{aligned}
\mathbf{G}_{kp} &= \left.\frac{\partial \mathbf{g}_{kp}(\mathbf{x}_k, \theta_i)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k-1|k-1}, \theta_i} \\
&= \begin{bmatrix} \frac{\partial a_{kp}}{\partial x_k(1)} p_{kp} r_{kp} + a_{kp} \frac{\partial p_{kp}}{\partial x_k(1)} r_{kp} + a_{kp} p_{kp} \frac{\partial r_{kp}}{\partial x_k(1)} \\ a_{kp} \frac{\partial p_{kp}}{\partial x_k(2)} r_{kp} + a_{kp} p_{kp} \frac{\partial r_{kp}}{\partial x_k(2)} \end{bmatrix}^T
\end{aligned}
\tag{7.19}
$$

The partial derivatives within the brackets are given by

$$\begin{aligned}
\frac{\partial a_{kp}}{\partial x_k(1)} &= \frac{-8A_{ip}}{c^2\hat{x}_k(1)^3} \\
\frac{\partial p_{kp}}{\partial x_k(1)} &= -\frac{\pi\hat{x}_k(2)}{t_w}\sin(2\pi\,\hat{x}_k(2)(k\,t_s-\hat{x}_k(1))/t_w)) \\
\frac{\partial r_{kp}}{\partial x_k(1)} &= \hat{x}_k(2)\,\omega_{cp}\sin\left[\hat{x}_k(2)(\omega_{cp}(k\,t_s-\hat{x}_k(1)))-\omega_{ip}k\,t_s\right] \\
\frac{\partial p_{kp}}{\partial x_k(2)} &= \frac{\pi\,(k\,t_s-\hat{x}_k(1))}{t_w}\sin(2\pi\,\hat{x}_k(2)(k\,t_s-\hat{x}_k(1))/t_w)) \\
\frac{\partial r_{kp}}{\partial x_k(2)} &= -\omega_{cp}\,(k\,t_s-\hat{x}_k(1)) \\
&\quad\times\sin\left[\hat{x}_k(2)(\omega_{cp}(k\,t_s-\hat{x}_k(1))-\omega_{ip}k\,t_s)\right]
\end{aligned} \tag{7.20}$$

The detection procedure consists of computing $\mathbf{g}_k(\mathbf{x}_k,\theta_i)$ and $\mathbf{G}_k(\mathbf{x}_k,\theta_i)$ for each value of $k\,t_s$ and for each model θ_i. For each model θ_i, if $x_k(1)\le k\,t_s < x_k(1)+t_w$ then $\mathbf{g}_k(\mathbf{x}_k,\theta_i)$ and $\mathbf{G}_k(\mathbf{x}_k,\theta_i)$ must be computed. The equations for the innovations $\tilde{\mathbf{z}}_{k|k-1,\theta_i}$ and covariance $S_{k|k-1,\theta_i}$ are given in [5.41, 5.42] for both the EKF and for the EHOF. Whenever the signal is assumed absent ($\hat{x}_k(1) < k\,t_s$, or $k\,t_s \ge \hat{x}_k(1)+t_w$), the innovations become

$$\tilde{\mathbf{z}}_{k|k-1,\theta_i} = z_k$$

$$S_{k|k-1,\theta_i} = R_k$$

With $\tilde{z}_{k|k-1,\theta_i}$ and $S_{k|k-1,\theta_i}$, $P(\theta_i|\mathbf{Z}_k)$ is computed using (7.9). The final state estimate is then computed using (7.10).

7.5 Joint Detection/Estimation of Time Delay

Under some conditions in which ν_k and τ_k are unknown it may be possible to obtain improved estimated of only one of these these state variables. For example,

if the pulse width is very short the minimum variance estimate of ν_k, which can be obtained from any unbiased estimator, may be larger than the initial Doppler variance. In this case Doppler estimation is redundant. This section addresses the model in which ν_k and τ_k are unknown, but only the time delay is to be estimated. Estimation of Doppler alone is addressed in the next section. The parameter vector θ_i is defined as before. To define hypothesis H_i, $\hat{\nu}_k$ is replaced by $\hat{\nu}_{m_0}$, that is, the Doppler parameter does not change from its initial estimate. Hypothesis H_i is now given by

$$H_i \quad : \quad z_{kp} = \begin{cases} v_k & k\,t_s < \hat{\tau}_k \\ g_{kp}(\hat{\tau}_k, \hat{\nu}_{m_0}) + v_k & \hat{\tau}_k \le k\,t_s < \hat{\tau}_k + t_w \\ v_k & k\,t_s \ge \hat{\tau}_k + t_w \end{cases} \tag{7.21}$$

with initial conditions

$$\begin{aligned} \hat{x}_{0|0,\theta_i} &= [\hat{\tau}_{n_0}, \hat{\nu}_{m_0}]^T \\ P_{0|0,\theta_i} &= \left[\mathrm{Var}[\tau_{n_0}]\right] \end{aligned} \tag{7.22}$$

where

$$g_{kp}(\hat{\tau}_k, \hat{\nu}_{m_0}) = a_{kp}(\hat{\tau}_k) p_{kp}(\hat{\tau}_k, \hat{\nu}_{m_0}) r_{kp}(\hat{\tau}_k, \hat{\nu}_{m_0}) \tag{7.23}$$

$$\begin{aligned} a_{kp}(\hat{\tau}_k) &= \frac{4A_{tp}}{(c\hat{\tau}_k)^2} \\ p_{kp}(\hat{\tau}_k, \hat{\nu}_{m_0}) &= 0.5\,(1 - \cos(2\pi\,\hat{\nu}_{m_0}(k\,t_s - \hat{\tau}_k)/t_w)) \\ r_{kp}(\hat{\tau}_k, \hat{\nu}_{m_0}) &= \cos\left[(\hat{\nu}_{m_0}(\omega_{cp}\,(k\,t_s - \hat{\tau}_k))) - \omega_{tp} k\,t_s\right] \end{aligned} \tag{7.24}$$

The state variable is $x_k = [\tau_k]$, and the Jacobian equations (7.19, 7.20) now contain only those terms that include partial derivatives of $x_k(1)$. The detection/estimation procedure is then the same as that described previously.

7.6 Joint Detection/Estimation of Doppler Shift

Consider the model in which ν_k and τ_k are unknown, but only the Doppler shift parameter is to be estimated. The parameter vector θ_i is defined as before. To define hypothesis H_i $\hat{\tau}_k$ is replaced by $\hat{\tau}_{n_0}$, that is, the time delay does not change from its initial estimate. Hypothesis H_i is now expressed as

$$H_i \quad : \quad z_{kp} = \begin{cases} v_k & k\,t_s < \hat{\tau}_{n_0} \\ g_{kp}(\hat{\tau}_{n_0}, \hat{\nu}_k) + v_k & \hat{\tau}_{n_0} \le k\,t_s < \hat{\tau}_{n_0} + t_w \\ v_k & k\,t_s \ge \hat{\tau}_{n_0} + t_w \end{cases} \tag{7.25}$$

with initial conditions

$$\begin{aligned} \hat{x}_{0|0,\theta_i} &= [\hat{\tau}_{n_0}, \hat{\nu}_{m_0}]^T \\ P_{0|0,\theta_i} &= \left[\mathrm{Var}[\tau_{n_0}]\right] \end{aligned} \tag{7.26}$$

where

$$g_{kp}(\hat{\tau}_{n_0}, \hat{\nu}_k) = a_{kp}(\hat{\tau}_{n_0}) p_{kp}(\hat{\tau}_{n_0}, \hat{\nu}_k) r_{kp}(\hat{\tau}_{n_0}, \hat{\nu}_k) \tag{7.27}$$

$$\begin{aligned} a_{kp}(\hat{\tau}_{n_0}) &= \frac{4A_{tp}}{(c\hat{\tau}_{n_0})^2} \\ p_{kp}(\hat{\tau}n_0, \hat{\nu}_k) &= 0.5\,(1 - \cos(2\pi\,\hat{\nu}_k(k\,t_s - \hat{\tau}_{n_0})/t_w)) \\ r_{kp}(\hat{\tau}_{n_0}, \hat{\nu}_k) &= \cos\left[(\hat{\nu}_k(\omega_{cp}\,(k\,t_s - \hat{\tau}_{n_0}))) - \omega_{tp}\,k\,t_s\right] \end{aligned} \tag{7.28}$$

The state variable is $x_k = [\nu_k]$, and the Jacobian equations (7.19, 7.20) now contain only those terms that include partial derivatives of $x_k(2)$. The detection/estimation procedure is described in section 7.4.

7.7 Detection Without Estimation

For the detection-only problem we use the initial estimates $\hat{\tau}_{n_0}$ and $\hat{\nu}_{m_0}$ in the measurement equation. The state estimates remain invariant. The hypothesis H_i that corresponds to parameter θ_i is defined as

$$H_i \quad : \quad z_k = \begin{cases} v_k & k\,t_s < \hat{\tau}_{n_0} \\ g_k(\hat{\tau}_{n_0}, \hat{\nu}_{m_0}) + v_k & \hat{\tau}_{n_0} \leq k\,t_s < \hat{\tau}_{n_0} + t_w \\ v_k & k\,t_s \geq \hat{\tau}_{n_0} + t_w \end{cases} \tag{7.29}$$

where

$$g_k(\hat{\tau}_{n_0}, \hat{\nu}_{m_0}) = a_k(\hat{\tau}_{n_0}) p_k(\hat{\tau}_{n_0}) r_k(\hat{\tau}_{n_0}, \hat{\nu}_{m_0}) \tag{7.30}$$

The procedure for determining the a *posteriori* probability is the same as that described in the previous section with the exception that the state variables are held constant at their initial estimates $\hat{\tau}_{n_0}$ and $\hat{\nu}_{n_0}$.

7.8 Experimental Evaluation

As a prelude to the experimental evaluation it is useful to discuss the minimum variance that can be obtained through the estimation of time delay and Doppler shift. Consider the measurement model for a single-frequency pulse in a rectangular window of size K, where K is the number of samples per pulse. This signal is expressed by

$$h_k = \sin(\nu\omega_c(kt_s - \tau)) \qquad 0 \leq k < K$$

If the signal h_k is received in additive white Gaussian noise with variance σ^2, the

Cramer Rao bound (Appendix A) is given by

$$\mathrm{Var}[\tau] \geq \frac{\sigma^2}{\sum_{k=0}^{K-1} E\left\{\left(\frac{\partial h_k}{\partial \tau}\right)^2\right\}} = \frac{1}{\mathrm{SNR}} \frac{1}{\nu^2 \omega_c^2 K} \tag{7.31}$$

$$\mathrm{Var}[\nu] \geq \frac{\sigma^2}{\sum_{k=0}^{K-1} E\left\{\left(\frac{\partial h_k}{\partial \nu}\right)^2\right\}} = \frac{1}{\mathrm{SNR}} \frac{1}{\omega_c^2 t_s^2 (\sum_{k=0}^{K-1} k^2)} \tag{7.32}$$

where SNR is the mean square signal amplitude divided by the white noise variance, i.e. $\mathrm{SNR} = A^2/(2\sigma^2)$. The delay variance is reduced by increasing the carrier frequency ω_c or by increasing the number of samples. Doppler variance is reduced by increasing the carrier frequency or by increasing the pulse width (i.e. by integrating over a longer time). The bound for time delay is achievable only if the initial uncertainty $\Delta\tau_0 < 1/(2f_c)$. Measurement of time delay is actually accomplished by measuring the phase of the received signal. Since the phase is periodic at a rate f_c, two time delay estimates separated by $1/f_c$ will give the same phase measurement for a single-frequency rectangular pulse. That is, an initial estimate $\hat{\tau} \geq \tau + 1/(2f_c)$ is more likely to converge to $\tau + 1/f_c$ than it is to τ. The situation can be improved somewhat by employing amplitude modulation or angle modulation. However, as shown in the following section, window functions such as the Hanning window do not help appreciably. Thus the variance of time delay error may be more accurately bounded by

$$\mathrm{Var}[\tau] \geq \begin{cases} \frac{1}{\mathrm{SNR}} \frac{1}{\nu^2 \omega_c^2 N} & \Delta\tau_0 < 1/(2f_c) \\ \mathrm{Var}[\tau_0] & \Delta\tau_0 \geq 1/(2f_c) \end{cases} \tag{7.33}$$

If the initial estimation error for time delay is not known to within $\Delta\tau = 1/(2f_c)$ then parameter estimators will not do any better than the initial estimates.

This is especially true whenever the SNR is large enough to nullify the effects of any pulse shaping. This limitation drives the requirement for the number of different filters needed for accurate delay estimation. If $\tilde{\tau}_0 = L\Delta\tau_0$ is the total width of the uniform distribution of the initial delay error, then the number of required filters is

$$L = 2f_c\tilde{\tau}_0 \tag{7.34}$$

For example assume a typical radar operating frequency of 10 GHz. For unambiguous range resolution the associated initial time delay uncertainty must be less than 1×10^{-10} seconds, corresponding to a range uncertainty, computed from equation (7.3), of 0.05 feet. It may be more appropriate to discuss time delay estimation in the context of communications signals where the operating frequencies are much lower and the pulse widths larger.

One method for dealing with this problem is to ensure that an initial estimate is within $\pm 1/(2f_c)$ of the actual time delay τ. This can be accomplished by segmenting the initial conditions and operating several estimators in parallel as described in Sections 7.2 - 7.6. This procedure is evaluated experimentally in the next section.

Another technique for time delay estimation discussed in the radar literature [78, pp.167-169] involves leading- and trailing-edge detection of the envelope of the received signal. The rise time t_R of the pulse is lower-bounded by the bandwidth f_B of the received signal with $t_R \approx 1/f_B$. The receiver includes a bandpass filter of width f_B, an envelope detector, and a threshold stage. For this type of receiver, the variance of the time delay estimate error is lower-bounded by [79, p. 299], [80,

pp. 400-404]

$$\mathrm{Var}[\tau] \geq \left[\frac{2E_r}{N_0}\right]^{-1} \frac{1}{4\pi^2 f_B^2}, \tag{7.35}$$

where $N_0/2$ is the noise power spectral intensity, and E_r is the received signal energy. The variance of the frequency estimation error, obtained from coherently processing the received signal, is bounded by

$$\mathrm{Var}[\omega_c] \geq \left[\frac{2\bar{E}_r}{N_0}\right]^{-1} \frac{1}{t_w^2}. \tag{7.36}$$

Noting that $\mathrm{Var}[\nu] = \mathrm{Var}[\omega_c]/\omega_c^2$, (7.36) is the continuous-time equivalent of (7.32). The variance in estimating Doppler shift is reduced by increasing the signal pulse width or by employing a larger carrier frequency. For envelope detection the variance in estimating time delay is reduced by increasing the signal bandwidth f_B. The ideal situation is to design the signal to obtain good estimation of both delay and Doppler. This is generally accomplished by employing amplitude and/or angle modulation on the pulse. The modulation is designed to produce large bandwidth (low $\mathrm{Var}[\tau]$), while a large t_w produces low $\mathrm{Var}[\omega_c]$.

Delay and Doppler estimation are addressed separately in the following experimental evaluation of the JD/E technique. Delay estimation is evaluated in section 7.7.1 for a signal with relatively small pulse width (large bandwidth). Doppler estimation is performed in section 7.7.2 for a signal with large pulse width.

7.8.1 Time Delay Estimation

Both single and double sensor models ($P = 1$, and $P = 2$) in (7.7) were selected for experimental evaluation. For this evaluation the sampling frequency was $f_s = 100 \times 10^6$ Hz, the pulse width was set to $12\,t_s$ and c, the speed of propagation, was 186000 miles/sec. For all tests, the nominal time delay and Doppler were

$\tau_{\text{nom}} = 0.000324$ and $(\nu_{\text{nom}} - 1) = 8.96 \times 10^{-7}$ respectively, corresponding to a target at a nominal range of 10 miles, traveling at 300 mph Doppler velocity.

It was assumed that the error in the time delay estimate was uniformly distributed at $\pm 3.5\, t_s$ about the nominal delay. The corresponding variance is then $(7\, t_s)^2/12$. The error in the Doppler estimate was assumed to be uniformly distributed at $\pm 7.47 \times 10^{-7}$ about the nominal Doppler. This corresponds to an error in the Doppler velocity of ± 250 mph. The corresponding variance is 1.85×10^{-13}.

7.8.1.1 Single Sensor Evaluation

It is noted that the model in (7.14) does not change appreciably for the range of values for ν_k specified in Section 7.8.1. That is, the magnitude of the partial derivatives with respect to $x_k(2)$ (Doppler shift) in (7.19) are much lower than those with respect to $x_k(1)$ (time delay). In fact it was found experimentally that the filter gain corresponding to the Doppler shift parameter was very small resulting in negligible change in this parameter from its initial estimate. For this reason the results presented for joint detection/estimation (JD/E) are shown for time delay estimation only. In this case the measurement model (7.14) becomes $z_{kp} = g_{kp}(\hat{\tau}_k, \nu_{\text{nom}}) + v_k$, for $\hat{\tau}_k \leq k\, t_s < \hat{\tau}_k + t_w$, with $\hat{x}_{0|0,\theta_i} = \hat{\tau}_{n_0}$, and $P_{0|0,\theta_i} = \text{Var}(\tau_{n_0})$. Thus, for the JD/E technique, $\hat{\nu}_k$ is held constant at its initial estimate ν_{nom}. For the single sensor evaluation the carrier frequency was $\omega_c = 2\pi * 10 \times 10^6$. The translation frequency was $\omega_t = 0$. Since the signal is oversampled ($f_s = 10 f_c$), it is not necessary to translate the signal.

The single sensor model was used to compare the use of multiple filters ($N = 7$) to a single filter ($N = 1$) for joint detection/estimation. With only one filter, $\hat{x}_{0|0,\theta_1} = \hat{\tau}_{\text{nom}}$, $P_{0|0,\theta_1} = (7\, t_s)^2/12$, as described previously. The initial estimates

of time delay for the multiple filter implementation are given by $\hat{\tau}_{n_0} = (n-4) * t_s + \tau_{\text{nom}}$, $n = 1, 2, \cdots 7$. Thus, the initial delay estimates were separated by t_s, with $\text{Var}(\tau_{n_0}) = t_s^2/12$, $\forall n$. The a *priori* probabilities are given by $P(\theta_n|\mathbf{Z}_0) = 1/N$, $1 \leq n \leq N$.

Figure 7.1 illustrates typical simulation results for the JD/E performance with a bank of seven filters ($N = 7$). This figure displays the received signal $\mathbf{z}_k$, the estimation error $\tilde{\tau}_{k|k,\theta_i}$, the error covariance $P_{k|k,\theta_i}$, and the a *posteriori* probability $p(\theta_i|\mathbf{Z}_k)$ for all models $\theta_i, i = 1, \cdots, 7$. The SNR was 10dB for this example. This figure shows that although the covariance converges, the estimation error does not converge to zero for all models. However, the weighting provided by a *posteriori* probability allows the proper model selection.

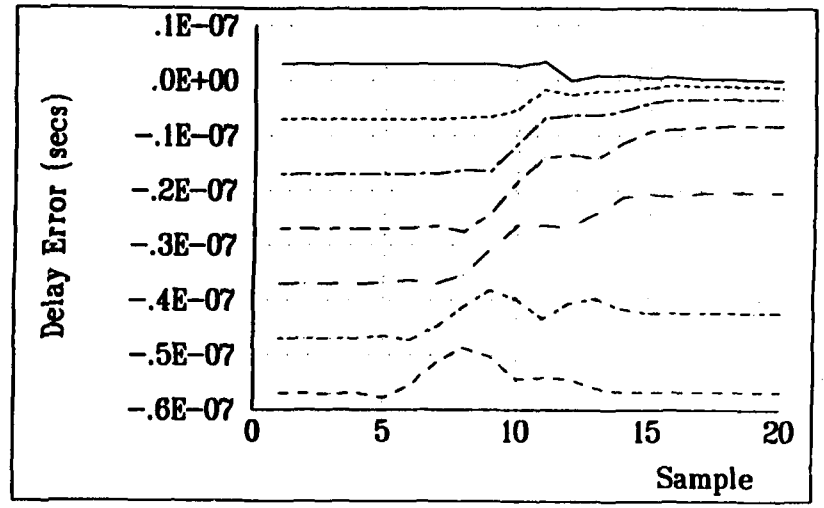
The Monte Carlo simulation results for JD/E with a single filter ($N = 1$) and a bank of seven filters ($N = 7$) are shown in Figure 7.2(a). In this figure the mean square error (MSE) of the estimation error in τ_k is shown as a function of SNR, where $\text{SNR} \equiv 10\log(E_s/\sigma_n^2)$, for $\tau_k \leq k\,t_s < \tau_k + t_w$, and E_s is the average received signal energy per sample. Each point on the graph represents the results of 500 simulation runs. Both the MAP and MMSE estimates are shown in Figure 7.2(a). The MAP and MMSE estimates are the same for $N = 1$. Also shown on this graph are the results for the detection-only (D-O) technique. The noise is Gaussian, and the EKF is used to perform estimation in the JD/E method. The JD/E ($N = 7$) implementation gives better results than the D-O method, particularly at higher SNR. This is expected since the filter in the JD/E method allows a considerable refinement estimates at higher SNR as compared to low SNR where the larger noise covariance restricts the filter gain. At -5 dB SNR the JD/E and D-O implementations perform identically. In general, the MMSE estimates are better than the MAP

estimates, particularly at low SNR's. The JD/E ($N = 1$) implementation gives the worst overall performance. The filter used in this implementation often converges to poor final estimates due to the tendency, mentioned previously, of time delay to converge to values that are separated from the actual time delay by multiples of $\pm 1/f_c$,

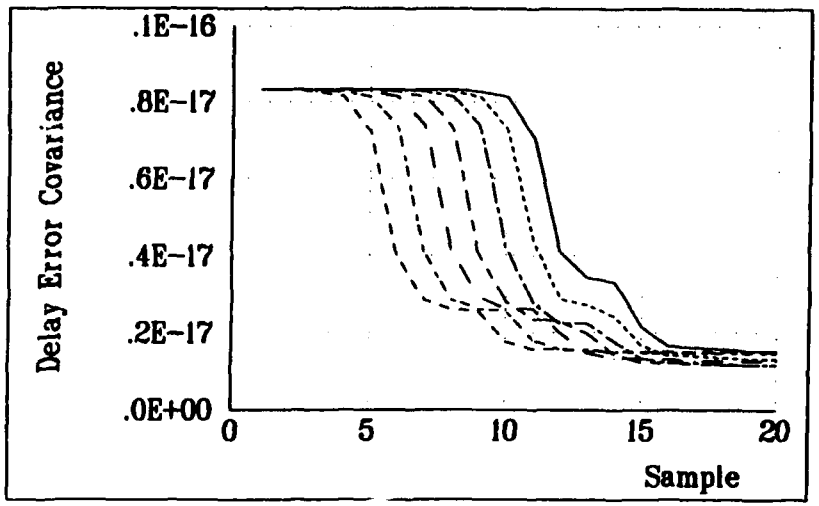
The importance of selecting an initial estimate within $\pm 1/(2f_c)$ is illustrated in Figure 7.3. This figure compares the JD/E MMSE (eqn 7.10) error distribution for $N = 1$ to that for $N = 7$. The distributions are formed from the results of 500 trials at each of the 6 SNR values $-5\text{dB}, 0\text{dB}, \cdots, 20\text{dB}$, giving 3000 total observations of the time delay estimation error. For the JD/E $N = 1$ case, where the initial estimation error is allowed to range between $\pm 3.5ts$ ($\pm 0.35/f_c$) for a single filter, a significant portion of the error distribution is centered around 1×10^7, of $1/f_c$. However, for the $N = 7$ case, in which the initial error distribution is segmented among the 7 filters, the entire distribution is centered around 0 error. In addition, the distribution around 0 error for $N = 7$ is tighter than the distribution around 0 for $N = 1$. This suggests that if the number of filters is chosen such that the initial estimation error of at least one of the filters is small in terms of $1/(2f_c)$, then the JD/E procedure can overcome the restriction on the initial estimation error imposed by (7.33).

The JD/E ($N = 7$) technique is evaluated in lognormal noise in Figure 7.2(b) for the single sensor model. The MMSE estimates of τ_k are shown in this figure for the EKF and for the EHOF. The EKF is evaluated in two configurations. In the first configuration, the Gaussian pdf is used to evaluate the detection statistic given by equation (7.9). In the second configuration, the lognormal pdf is used. The EHOF is evaluated using the lognormal pdf only. The EKF in the second configuration and

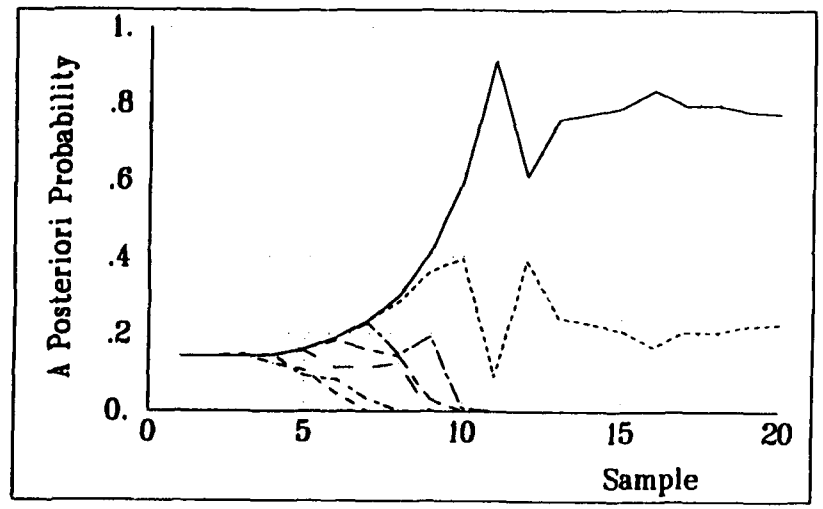
(a) Received Signal $\mathbf{z}_k$



(c) Time Delay Estimation Error $\tilde{\tau}_{k|k,\theta_i}$

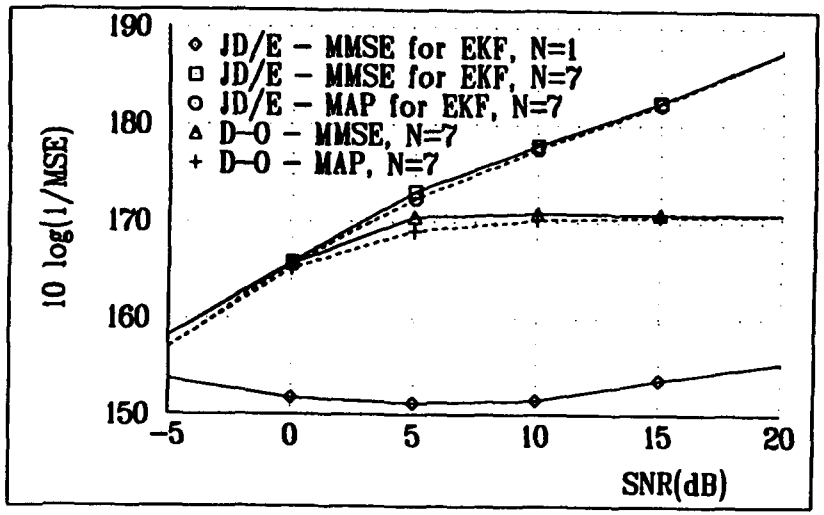


(b) Error Covariance $P_{k|k,\theta_i}$

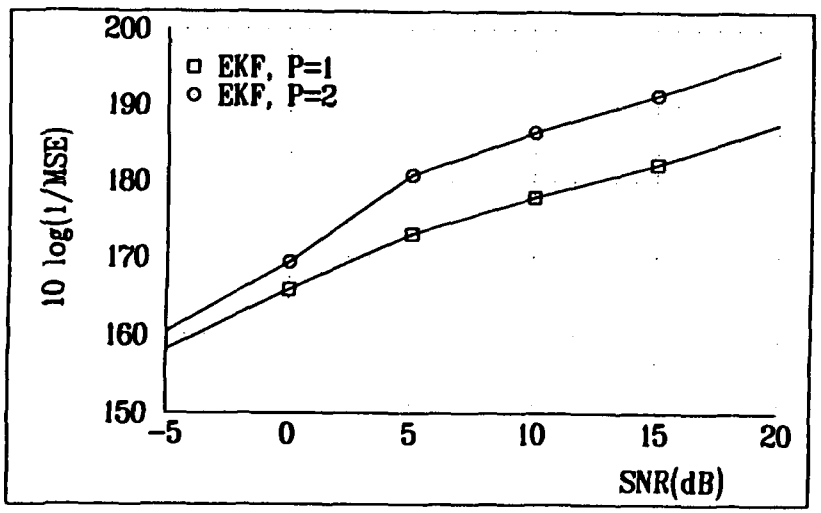


(d) *A Posteriori* Probability $p(\theta_i|\mathbf{Z}_k)$

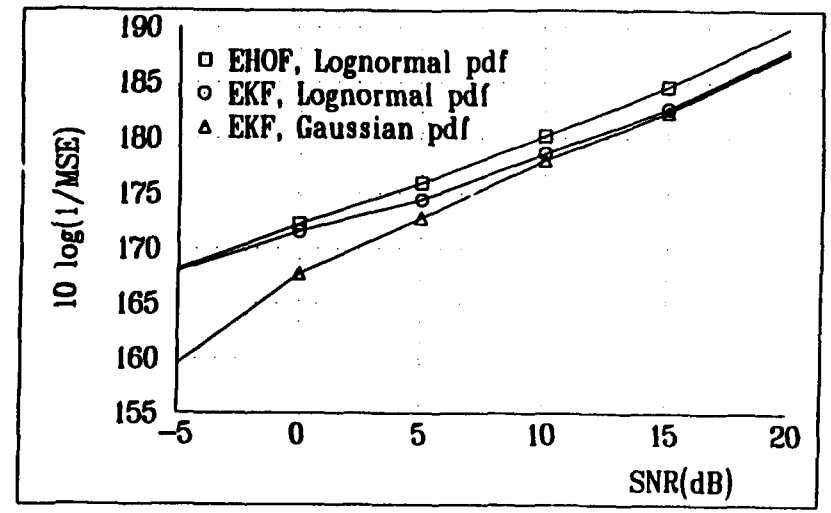
Figure 7.1 Typical JD/E Results for $N = 7$, 10dB SNR



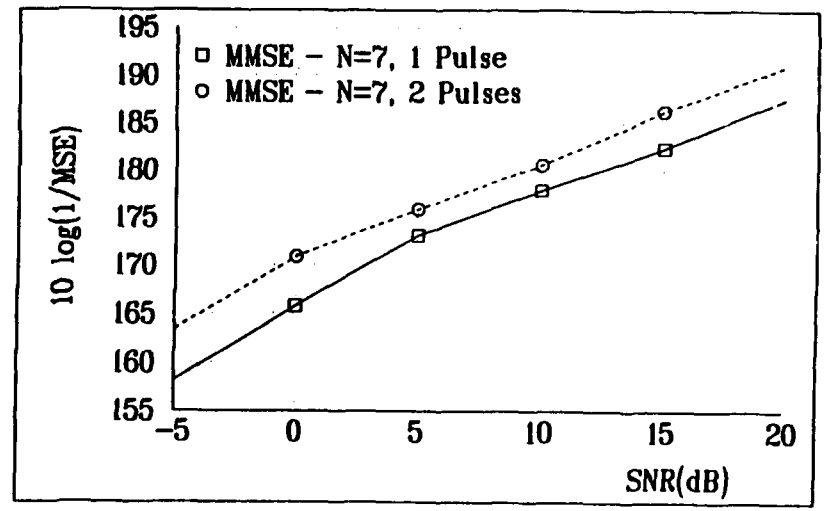
(a) Error in $\hat{\tau}_k$, Gaussian Noise, $P = 1$



(b) JD/E - MMSE for $\hat{\tau}_k$, $N = 7$, Multiple Sensor



(c) JD/E - MMSE for $\hat{\tau}_k$, Lognormal Noise, $N = 7$, $P = 1$



(d) JD/E - MMSE for $\hat{\tau}_k$, $N = 7$, Multiple Pulses

Figure 7.2 JD/E Performance for Time Delay Estimation

the EHOF give very similar results at low SNR. However, at high SNR the EHOF outperforms the EKF. When the Gaussian pdf is used in conjunction with the EKF to perform detection, the results are much worse than when the proper lognormal pdf is used. This advantage is particularly evident at low SNR's.

7.8.1.2 Double Sensor Evaluation

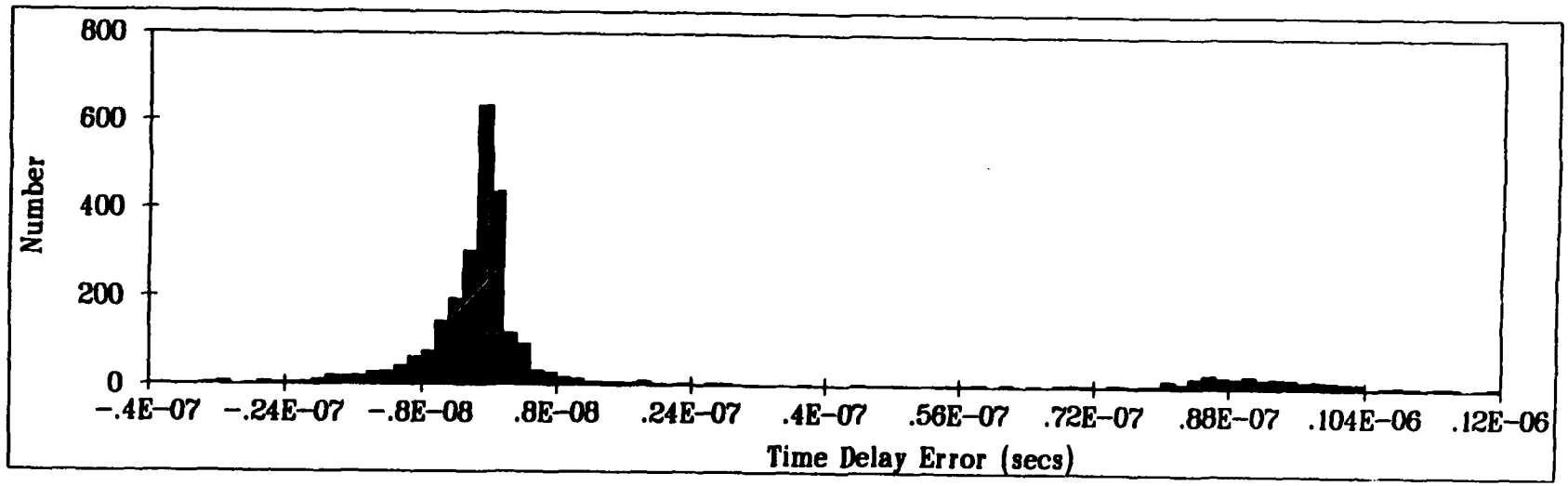
In the multiple sensor case ($P > 1$), the sensors may have different carrier frequencies (ω_{c_p}), and different translation frequencies (ω_{t_p}). A two-sensor ($P = 2$) model was evaluated in which $\omega_{c_1} = 2\pi * 10 \times 10^6$, $\omega_{c_2} = 2\pi * 30 \times 10^6$, and $\omega_{it_1} = \omega_{t_2} = 0$. The MMSE results of this evaluation for JD/E ($N = 7$) are given in Figure 7.2(c). The single-sensor ($P = 1$) MMSE results are also shown in this figure. This figure illustrates the distinct advantage of centralized fusion for JD/E.

7.8.1.3 Multiple Pulse Processing

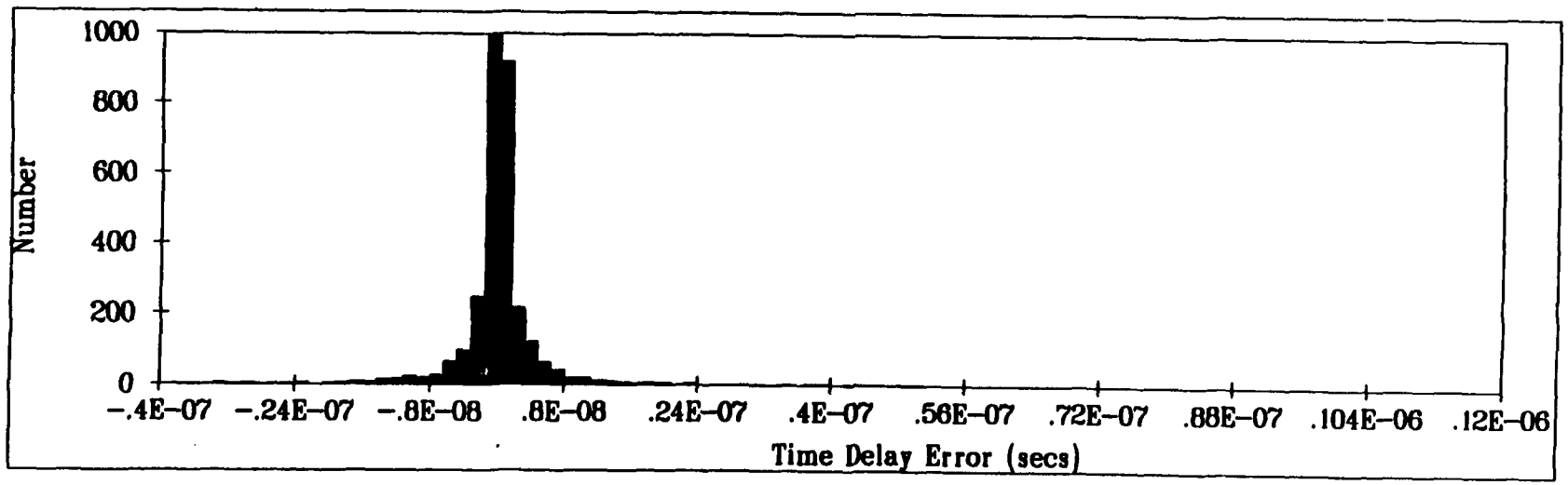
The results of processing two pulses are given in Figure 7.2(d). The EKF and EHOF are configured such that the initial error covariance is reset at the beginning of each pulse. The rationale for this, as discussed in Chapter 4, is to re-excite the system. This helps to allow poor estimates to possibly converge to smaller errors, and as shown in Chapter 4 it does not significantly effect those estimates that have already converged close to the actual state value. Figure 7.1(d) shows an improvement of about 3 dB for the two pulse estimate over the single pulse estimate. This improvement is supported by (7.31).

7.8.2 Doppler Estimation

Both single and double sensor models ($P = 1$, and $P = 2$) in (7.7) were selected for experimental evaluation for Doppler shift estimation. For this evalua-



(a) Time Delay Error Distribution for $N = 1$



(b) Time Delay Error Distribution for $N = 7$

Figure 7.3 Distribution of JD/E MMSE Time Delay Errors for $N = 1$, $N = 7$

tion, the pulse width was set to $24\,t_s$ and, c, the speed of propagation, was 186000 miles/sec. For all tests, the nominal time delay and Doppler were $\tau_{\text{nom}} = 0.000324$ and $(\nu_{\text{nom}} - 1) = 8.96 \times 10^{-7}$ respectively, corresponding to a target at a nominal range of 10 miles, traveling at 300 mph Doppler velocity. The sampling frequency was set to $f_s = 4 \times 10^3$ Hz. The error in the Doppler estimate was assumed to be uniformly distributed at $\pm 7.47 \times 10^{-7}$ about the nominal Doppler. This corresponds to an error in the Doppler velocity of ± 250 mph. The corresponding variance is 1.85×10^{-13}. It is observed from (7.32) that the Doppler error variance can be decreased by increasing the pulse width or by increasing the carrier frequency ω_c. This is the reason for the large pulse width (6 msec) for Doppler estimation versus the relatively small pulse width (0.12μsec) used for time delay estimation.

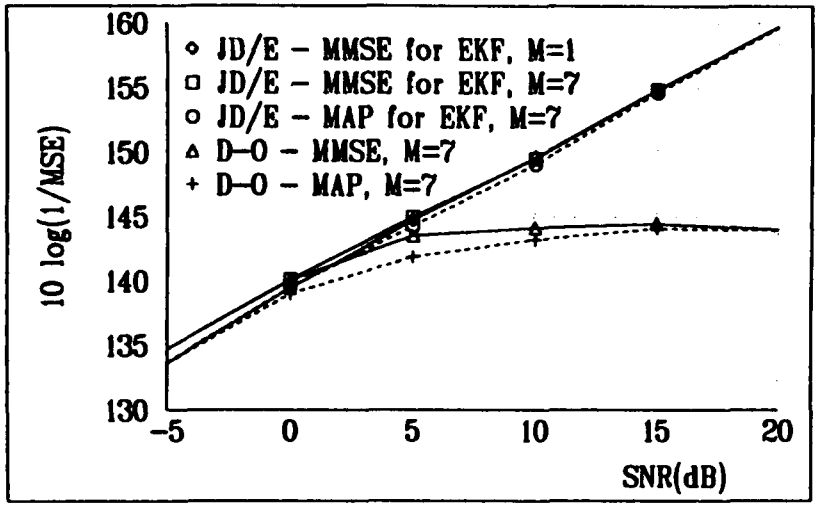
7.8.2.1 Single Sensor Evaluation

For Doppler-only estimation the measurement model (7.14) becomes $z_{k_p} = g_{k_p}(\hat{\tau}_{\text{nom}}, \nu_k) + v_k$, for $\hat{\tau}_{\text{nom}} \leq k\,t_s < \hat{\tau}_{\text{nom}} + t_w$, with $\hat{x}_{0|0,\theta_i} = \hat{\nu}_{m_0}$, and $P_{0|0,\theta_i} = \text{Var}(\nu_{m_0})$. Thus, for the JD/E technique, $\hat{\tau}_k$ is held constant at its initial estimate τ_{nom}. For the single sensor case the carrier and translation frequencies were $\omega_c = 2\pi * 100 \times 10^6$, and $\omega_t = 2\pi * 99.9975 \times 10^6$.

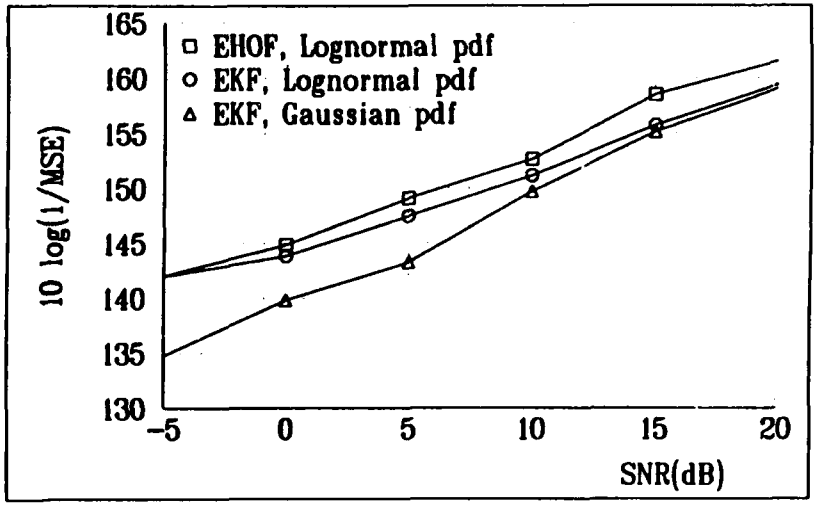
The single sensor model was used to compare the use of seven Doppler filters ($M = 7$) to a single filter ($M = 1$) for joint detection/estimation. With only one filter, $\hat{x}_{0|0,\theta_1} = \hat{\tau}_{\text{nom}}$, $P_{0|0,\theta_1} = (2\nu_{max})^2/12$, where ν_{max} is the maximum initial Doppler shift excursion due to the maximum Doppler velocity of 250 mph. $\nu_{max} = (2 * 250)/(3600 * c) = 7.46 \times 10^{-7}$. The initial estimates of Doppler for the multiple filter implementation are given by $\hat{\nu}_{m_0} = (m - 4) * \Delta\nu + \nu_{\text{nom}}$, $m = 1, 2, \cdots 7$, where $\Delta\nu = 2\nu_{max}/7$. The initial variance for each of the 7 filters is

$\mathrm{Var}(\nu_{m_0}) = (\Delta\nu)^2/12$, $\forall m$. The *a priori* probabilities are given by $P(\theta_m|\mathbf{Z}_0) = 1/M$, $1 \leq m \leq N$. The Monte Carlo simulation results for JD/E for $M = 1$ and $M = 7$ are shown in Figure 7.4(a). In this figure the mean square error (MSE) of the estimation error in ν_k is shown as a function of SNR. Both the MAP and MMSE estimates are shown in Figure 7.4(a). The MAP and MMSE estimates are the same for $M = 1$. Also shown on this graph are the results for the detection-only (D-O) technique. The JD/E ($M = 7$) results are essentially the same as the JD/E ($M = 1$) results. Thus, in this case there is no advantage in using more than one filter to estimate the Doppler shift. (This is in contrast to the time delay estimation results shown in Figure 7.2(a), in which the JD/E $N = 7$ performance was much better than that for JD/E $N = 1$.) The JD/E ($M = 7$) implementation gives better results than the D-O method, particularly at higher SNR. This is expected since the filter in the JD/E method allows a considerable refinement estimates at higher SNR as compared to low SNR where the larger noise covariance restricts the filter gain. At -5 dB SNR the JD/E and D-O implementations perform identically. In general, the MMSE estimates are better than the MAP estimates, particularly at low SNR's.

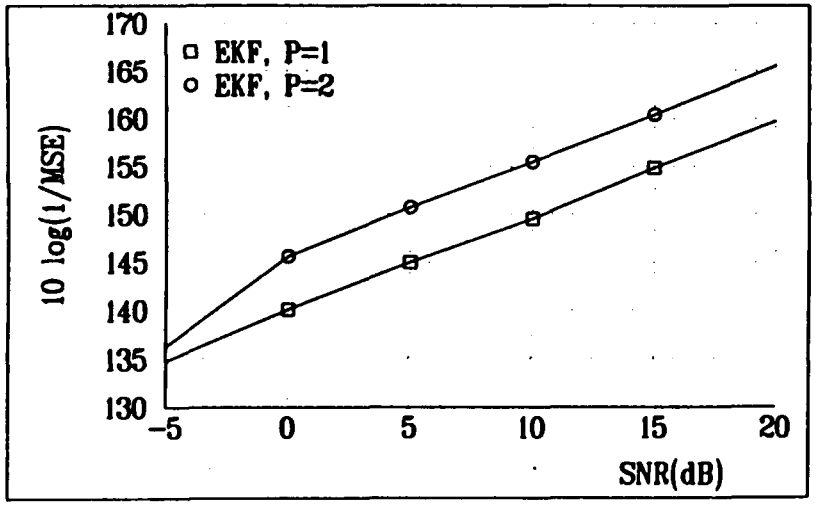
The JD/E ($M = 7$) technique is evaluated in lognormal noise in Figure 7.4(b) for the single sensor model. The MMSE estimates of ν_k are shown in this figure for the EKF and for the EHOF. The EKF is evaluated in two configurations. In the first configuration the Gaussian pdf is used to evaluate the detection statistic given by equation (7.9). In the second configuration the lognormal pdf is used. The EHOF is evaluated using the lognormal pdf only.



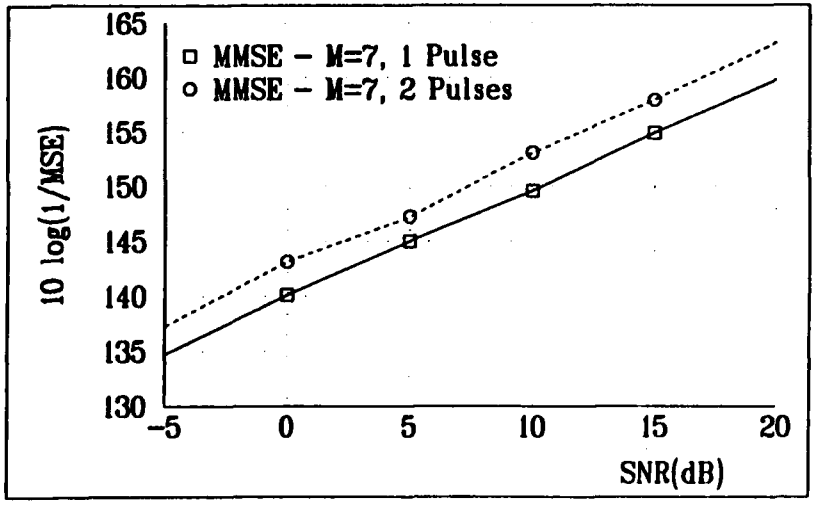
(a) Error in $\hat{\nu}_k$, Gaussian Noise, $P = 1$



(c) JD/E - MMSE for $\hat{\nu}_k$, Lognormal Noise, $N = 7$, $P = 1$



(b) JD/E - MMSE for $\hat{\nu}_k$, $N = 7$, Multiple Sensor



(d) JD/E - MMSE for $\hat{\nu}_k$, $N = 7$, Multiple Pulses

Figure 7.4 JD/E Performance for Doppler Shift Estimation

7.8.2.2 Double Sensor Evaluation

In the multiple sensor case ($P > 1$) the sensors may have different carrier frequencies (ω_{c_p}), and different translation frequencies (ω_{t_p}). A two-sensor ($P = 2$) model was evaluated in which $\omega_{c_1} = 2\pi * 100 \times 10^6$, $\omega_{t_1} = 2\pi * 99.9975 \times 10^6$, $\omega_{c_2} = 2\pi * 200 \times 10^6$, and $\omega_{t_2} = 2\pi * 199.9975 \times 10^6$. The MMSE results of this evaluation for JD/E ($M = 7$) are given in Figure 7.4(c). The single-sensor ($P = 1$) MMSE results are also shown in this figure. This figure illustrates the distinct advantage of centralized fusion for JD/E.

7.8.2.3 Multiple Pulse Processing

The double pulse model is compared to the single pulse model in Figure 7.4(d). Recall from (7.36) that the variance in frequency (and Doppler) estimates is a function of the inverse pulse width squared. However, processing two pulses does not give the same advantage as processing an equivalent pulse of size $2t_w$. As shown in Figure 7.4(d) the advantage is approximately 3 dB - the same as for time delay estimation (Figure 7.2(d)).

7.9 Conclusion

The space-time modeling of the signal returns as described in (7.14) has been used in conjunction with nonlinear filters to design a new adaptive sensor processor. Simulation results show excellent detection capabilities and excellent resolution in target parameter estimation for both single and multiple sensor data. With the excellent detectability, fine parameter resolution, and automatic data referencing, this approach presents a very competitive design for target detection and parameter estimation.

The most significant result from the implementation of the JD/E technique for time delay estimation is that the requirement that the initial estimation error $\Delta\tau < 1/(2f_c)$ can be relaxed by implementing several parallel filters.

Chapter 8

Multisensor Detection and Signal Parameter Estimation

This chapter addresses the problem of multi sensor detection and high resolution signal parameter estimation using joint maximum a posteriori detection and high order nonlinear filtering techniques. The specific problem addressed is that of two spatially separated sensors that employ active echo processing to estimate the parameters of a target. The geometric area of coverage of the two sensors is permitted to overlap. In the overlap region the estimates from the two sensors are combined to produce improved estimates over the single sensor estimates.

The problem is approached using joint detection/estimation techniques. Several hypotheses are postulated for detection. Each hypothesis corresponds to the ability of each sensor to detect the target in its area of coverage. The a *priori* probabilities of each decision are based on the area of coverage of the two sensors. For each hypothesis, a high order filter recursively estimates time delay, Doppler shift and geometric angle to the target from processing the returns of the transmitted signal from each sensor. These estimates are in turn used to estimate target position and velocity. For each of these hypotheses, another set of parallel filters is used to obtain more accurate estimates of signal parameters and to account for the stability problems that result from the first order Taylor series expansion used in the nonlinear filtering algorithms. This is accomplished by operating a separate filter for each of several different initial time delay estimates of the return signal. The maximum likelihood estimate for a given hypothesis is then determined as a weighted sum of the estimates from each of the local hypotheses, with the a *posteriori* probability being used as the weighting function. It is assumed that the signals are imbedded

in Gaussian noise, and clutter. The clutter is treated as non-Gaussian noise with a lognormal or Weibull distribution.

Consider the situation of two spatially separated sensors, $s1$ and $s2$. Each of the two sensors attempts to detect and track objects coming into its respective area of coverage. For a valid data fusion scenario, the coverage of the two sensors is assumed to overlap in space, but not entirely. The sensor geometry is shown in Figure 8.1. In the overlap region the data received by the two sensors can be combined to get a more accurate estimate of target parameters or to estimate parameters that cannot be estimated with one sensor alone. In the overlap region the estimates from the individual sensors are combined to form improved target parameter estimates. We consider the case where each of the sensors may have different types of tracking devices such as optical trackers, various types of radars, etc. It is assumed that these sensors transmit a signal and process the echo returned from that signal. It is assumed that the signals are corrupted by additive Gaussian noise due to thermal effects within the receiver, and by clutter which may be due to non-Gaussian distortion such as sea clutter or other multipath spreading. The amplitude of sea clutter is characterized by statistical fluctuations which may be described in terms of a probability density function. Typical distributions used to model this distortion include the Rayleigh, Weibull or lognormal distributions [81, pp. 478-479]. In general, the Gaussian noise introduced into the receiver is uncorrelated between the two sensors.

8.1 System Model

Assume that each sensor consists of a phased array or other sensing device that can produce target angle estimates along with estimates of time delay and

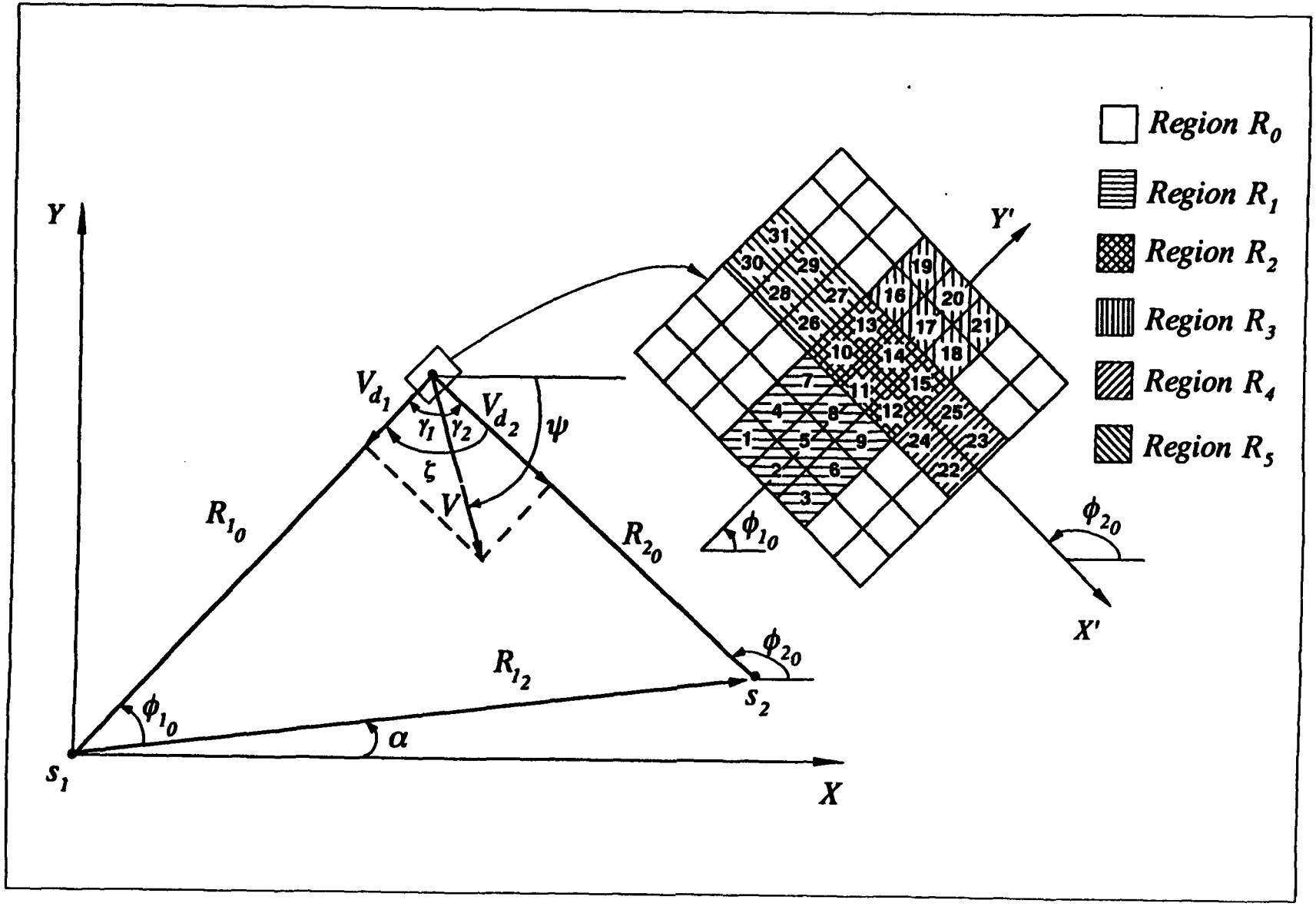


Figure 8.1 Sensor/Target Geometry for Multisensor Fusion

Doppler shift. It is assumed that there are two separate measurements taken at each sensor - one measurement at each of the offset phase centers. The received signal at the i^{th} sensor may be described by

$$\mathbf{z}_{i_k} = \mathbf{h}_{i_k} + \mathbf{u}_{i_k} + \mathbf{v}_{i_k} \tag{8.1}$$

where $\mathbf{h}_{i_k}$ represents the received signal, $\mathbf{u}_{i_k}$ is the clutter, and $\mathbf{v}_{i_k}$ is the Gaussian noise at the k^{th} sampling interval. Since there are two measurements, the received signal can be more explicitly stated as

$$\begin{bmatrix} z_{i1_k} \\ z_{i2_k} \end{bmatrix} = \begin{bmatrix} h_{i1_k} \\ h_{i2_k} \end{bmatrix} + \begin{bmatrix} u_{i1_k} \\ u_{i2_k} \end{bmatrix} + \begin{bmatrix} v_{i1_k} \\ v_{i2_k} \end{bmatrix} \tag{8.2}$$

The received signal vector $\mathbf{h}_{i_k}$ at sensor i can be described by the following model.

$$\begin{bmatrix} h_{i1_k} \\ h_{i2_k} \end{bmatrix} = \begin{bmatrix} a(kt_s - \tau_{i1} + \tau_{i2}/2)p(kt_s - \tau_{i1} + \tau_{i2}/2)\sin(\nu_i(\omega_i(kt_s - \tau_{i1} + \tau_{i2}/2))) \\ a(kt_s - \tau_{i1} - \tau_{i2}/2)p(kt_s - \tau_{i1} - \tau_{i2}/2)\sin(\nu_i(\omega_i(kt_s - \tau_{i1} - \tau_{i2}/2))) \end{bmatrix} \tag{8.3}$$

where $a(.)$ is the amplitude, $p(.)$ is the pulse shaping function, and t_s is the sample time ($t_s = 1/f_s$). The delay τ_{i1} is the round-trip propagation time from the center of the sensor to the target and back to the sensor. Referring to Figure 8.2, this is the time for the signal to travel from point P_i to O and back to point P_i. From τ_{i1} the range to the target can be determined using the relationship

$$R_i = \frac{\tau_{i1}}{2c} \tag{8.4}$$

where c is the speed of propagation. The delay τ_{i2} is the difference in time for the signal to reach from point P_{i1} to point P_{i2}. The difference in the propagation distance is given by $c\tau_{i2}$. The differential angle $\Delta\phi_i$ to the target from sensor i, which represents the difference between the sensor pointing angle ϕ_{i_0} and the actual

target angle ϕ_i, is then

$$\Delta\phi_i = \sin^{-1}\left(\frac{c\tau_{i2}}{D_i}\right)$$
$$\phi_i = \phi_{i_0} + \Delta\phi_i \tag{8.5}$$

where D_i is the distance between the two offset phase centers in the phased array for sensor i. The initial estimates of τ_{i2} are based on the geometric relationship shown in Figure 8.2. This figure shows that the geometric angle $\Delta\phi_i$ is given by

$$\sin(\Delta\phi_i) = \frac{\Delta r_i}{R_i} = \frac{c\tau_{i2}}{D_i} \tag{8.6}$$

The function $p(.)$ in (8.3) represents the pulse shaping function and is generally designed to limit the signal bandwidth at the expense of widening the main lobe of the function in the frequency domain. Several possible pulse shapes and their spectral characteristics are given by Harris [82]. It is assumed that the signal is attenuated by spherical spreading loss such that the received amplitude $a(.)$ is related to the transmitted amplitude A through the relation

$$a(kt_s - (\tau_{i1} \pm \tau_{i2}/2)) = \frac{4A}{(c(\tau_{i1} \pm \tau_{i2}/2))^2} \tag{8.7}$$

For constant receiver noise power σ_g^2, the signal to white noise ratio at the receiver is given by

$$\mathrm{SNR}_i = \frac{E_p a(kt_s - (\tau_{i1} \pm \tau_{i2}/2))^2}{2\sigma_g^2} = \frac{8E_pA^2}{c^2(\tau_{i1} \pm \tau_{i2}/2)^2\sigma_g^2} \tag{8.8}$$

where E_p is the average pulse energy per sample. Given that the transmitted amplitude A and the carrier frequency ω_i are known, then the unknown delays τ_{i1} and τ_{i2}, and the Doppler shift parameter ν_i must be estimated. The Doppler velocity V_{d_i}, which is the projection of the target velocity along the line of sight from sensor i to the target and is given by $V_{d_i} = |\mathbf{V}|\cos(\gamma_i)$, where $\mathbf{V}$ is the target velocity vector.

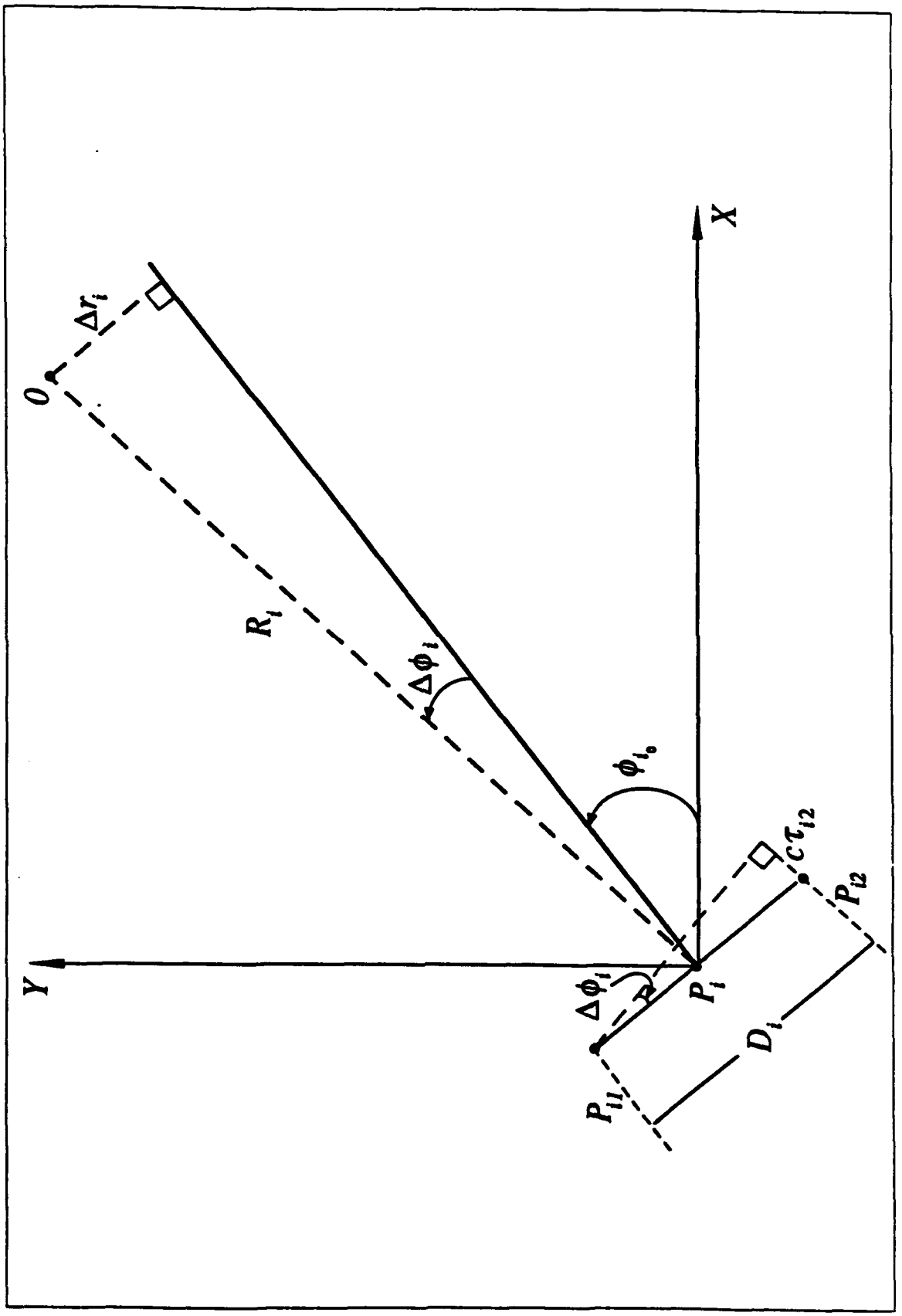


Figure 8.2 Single Sensor Intercept Geometry

V_{d_i} is related to ν_i through the relation

$$V_{d_i} = \frac{\nu_i - 1}{2c} \tag{8.9}$$

The estimation accuracy and the number of target motion parameters that can be estimated are a function of the position of the target within the coverage area of each sensor. If the target is located in a region covered only by a single sensor then estimates of the Doppler shift and the two time delays from this single sensor can be used to estimate only the target position and Doppler velocity V_{d_i} for that sensor. If the target is in the overlap region then estimates of the Doppler shift and the two delays from each sensor can be used to obtain a more accurate estimate of target position in addition to estimating the complete target velocity vector. The models for these two situations are developed next.

8.1.1 Single Observer Model

Using estimates of τ_{i1}, τ_{i2} and ν_i from one sensor the target position and Doppler velocity can be estimated through the relations (8.4, 8.5, and 8.9). Define the state variable vector for sensor i as

$$\mathbf{x}_{i_k} = \left[\tau_{i1_k}\ \tau_{i2_k}\ \nu_{i_k}\right]^T$$

It is assumed that the state does not change while the pulse is being reflected from it. Therefore the process equation is not necessary; that is, the state transition matrix is unity and there is no process noise. In terms of the state variables the received signal at transmitter i is

$$\mathbf{h}_{i_k} = \begin{bmatrix} a_{i1_k}(\mathbf{x}_{i_k}) p_{i1_k}(\mathbf{x}_{i_k}) r_{i1_k}(\mathbf{x}_{i_k}) \\ a_{i2_k}(\mathbf{x}_{i_k}) p_{i2_k}(\mathbf{x}_{i_k}) r_{i2_k}(\mathbf{x}_{i_k}) \end{bmatrix} \tag{8.10}$$

where

$$
\begin{aligned}
a_{i1_k}(\mathbf{x}_{i_k}) &= \frac{4A}{(c(x_{i_k}(1)+x_{i_k}(2)/2))^2} \\
a_{i2_k}(\mathbf{x}_{i_k}) &= \frac{4A}{(c(x_{i_k}(1)-x_{i_k}(2)/2))^2} \\
p_{i1_k}(\mathbf{x}_{i_k}) &= 0.5*(1-\cos(2\pi x_{i_k}(3)(kt_s - x_{i_k}(1)+x_{i_k}(2)/2)/t_{w_i})) \\
p_{i2_k}(\mathbf{x}_{i_k}) &= 0.5*(1-\cos(2\pi x_{i_k}(3)(kt_s - x_{i_k}(1)-x_{i_k}(2)/2)/t_{w_i})) \\
r_{i1_k} &= \cos(x_{i_k}(3)(\omega_i(kt_s - x_{i_k}(1)+x_{i_k}(2)/2))) \\
r_{i2_k} &= \cos(x_{i_k}(3)(\omega_i(kt_s - x_{i_k}(1)-x_{i_k}(2)/2)))
\end{aligned}
\tag{8.11}
$$

The definition of $p_{i1_k}(\mathbf{x}_{i_k})$ given above represents the Hanning pulse type with pulse width t_{w_i}. The filter equations require the derivative of the signal model with respect to the state. This derivative is given by

$$
\frac{\partial \mathbf{h}_{i_k}}{\partial \mathbf{x}_{i_k}} = \begin{bmatrix} \frac{\partial a_{i1_k}}{\partial x_{i_k}} p_{i1_k} r_{i1_k} + a_{i1_k} \frac{\partial p_{i1_k}}{\partial x_{i_k}} r_{i1_k} + a_{i1_k} p_{i1_k} \frac{\partial r_{i1_k}}{\partial x_{i_k}} \\ \frac{\partial a_{i2_k}}{\partial x_{i_k}} p_{i2_k} r_{i2_k} + a_{i2_k} \frac{\partial p_{i2_k}}{\partial x_{i_k}} r_{i2_k} + a_{i2_k} p_{i2_k} \frac{\partial r_{i2_k}}{\partial x_{i_k}} \end{bmatrix} \tag{8.12}
$$

where

$$\frac{\partial a_{ij_k}}{\partial x_{i_k}(1)} = \frac{-8A}{c^2\left((x_{i_k}(1) - \kappa_j x_{i_k}(2)/2)\right)^3}$$

$$\frac{\partial a_{ij_k}}{\partial x_{i_k}(2)} = \frac{\kappa_j 4A}{c^2\left((x_{i_k}(1) - \kappa_j x_{i_k}(2)/2)\right)^3}$$

$$\frac{\partial p_{ij_k}}{\partial x_{i_k}(1)} = -(\pi x_{i_k}(3)/t_{w_i})\sin(2\pi x_{i_k}(3)(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)/t_{w_i})$$

$$\frac{\partial p_{ij_k}}{\partial x_{i_k}(2)} = 0.5\kappa_j \pi x_{i_k}(3)/t_{w_i}\sin(2\pi x_{i_k}(3)(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)/t_{w_i})$$

$$\frac{\partial p_{ij_k}}{\partial x_{i_k}(3)} = (\pi(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)/t_{w_i}) \qquad (8.13)$$

$$\sin(2\pi x_{i_k}(3)(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)/t_{w_i})$$

$$\frac{\partial r_{ij_k}}{\partial x_{i_k}(1)} = +x_{i_k}(3)\omega_i\sin(x_{i_k}(3)(\omega_i(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)))$$

$$\frac{\partial r_{ij_k}}{\partial x_{i_k}(2)} = -0.5\kappa_j x_{i_k}(3)\omega_i\sin(x_{i_k}(3)(\omega_i(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)))$$

$$\frac{\partial r_{ij_k}}{\partial x_{i_k}(3)} = -\omega_i(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)$$

$$\sin(x_{i_k}(3)(\omega_i(kt_s - x_{i_k}(1) + \kappa_j x_{i_k}(2)/2)))$$

for $j = 1, 2$. $\kappa_j = +1$ whenever $j = 1$. $\kappa_j = -1$ whenever $j = 2$.

It is assumed that the error in the initial conditions is not correlated with the measurement noise and that the Gaussian noise is not correlated with the clutter. Thus the measurement noise covariance is given by

$$\begin{aligned} R_{i_k}^{(2)} &= E[(\mathbf{u}_{i_k} + \mathbf{v}_{i_k})(\mathbf{u}_{i_k} + \mathbf{v}_{i_k})^T] \\ &= E[\mathbf{u}_{i_k}\mathbf{u}_{i_k}^T] + E[\mathbf{v}_{i_k}\mathbf{v}_{i_k})^T] \\ &= \sigma_c^2 \begin{bmatrix} 1 & \rho_i \\ \rho_i & 1 \end{bmatrix} + \sigma_g^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \qquad (8.14)$$

where it is assumed that the Gaussian measurement noise and the clutter are uncorrelated, and that the clutter has correlation coefficient ρ_i between the two offset phase center beams. The third moment of the measurement noise is

$$R_{i_k}^{(3)} = E[(\mathbf{u}_{i_k} \otimes \mathbf{u}_{i_k})\mathbf{u}_{i_k}^T] \tag{8.15}$$

which contains components only due to the clutter since the Gaussian noise has a symmetrical density function. The fourth moment of the measurement noise is

$$R_{i_k}^{(4)} = E[(\mathbf{u}_{i_k} \otimes \mathbf{u}_{i_k})(\mathbf{u}_{i_k} \otimes \mathbf{u}_{i_k})^T] + E[(\mathbf{v}_{i_k} \otimes \mathbf{v}_{i_k})(\mathbf{v}_{i_k} \otimes \mathbf{v}_{i_k})^T] \tag{8.16}$$

8.1.2 Double Observer Model

When information is available from two sensors, that is, whenever the target is in the overlap region, and the target is illuminated simultaneously by the two radars, the Doppler and time delay estimates from each sensor can be combined to obtain a better estimate of target position and velocity.

Let X' and Y' denote the directions of a local coordinate system as shown in the insert in Figure 8.1. Let ϕ_{1_0} and ϕ_{2_0}, the pointing angles of the two sensors, be chosen such that $\phi_{2_0} - \phi_{1_0} = 90\,\text{deg}$. In this case the direction X' points directly along the line of sight (LOS) of s_2, and perpendicular to the LOS of s_1. Likewise, Y' points directly along the LOS of s_1 and perpendicular to the LOS of s_2. X' is the in-track direction for s_1 and the cross-track direction for s_2. Y' is it in-track direction for s_2 and the cross-track direction for s_1. For small angles $\Delta\phi_i$ such that

$\sin(\Delta\phi_i \approx 0)$, the position estimates in the X', Y' coordinate system are given by

$$\begin{aligned} \hat{O}_{x'} &= -(c\hat{\tau}_{21}/2 - R_{2_0}) \\ &= R_{1_0} c\hat{\tau}_{12}/D_1 \\ \hat{O}_{y'} &= c\hat{\tau}_{11}/2 - R_{1_0} \\ &= R_{2_0} c\hat{\tau}_{22}/D_2 \end{aligned} \tag{8.17}$$

where R_{i_0} is the nominal range from sensor i to the center of the insert in Figure 8.1. The associated position error variances are given by

$$\begin{aligned} \sigma^2_{x'_1} &= R^2_{1_0} c^2 \mathrm{Var}[\tau_{12}]/D_1^2 \\ \sigma^2_{x'_2} &= c^2 \mathrm{Var}[\tau_{21}]/4 \\ \sigma^2_{y'_1} &= c^2 \mathrm{Var}[\tau_{11}]/4 \\ \sigma^2_{y'_2} &= R^2_{2_0} c^2 \mathrm{Var}[\tau_{22}]/D_2^2 \end{aligned} \tag{8.18}$$

If it is assumed that the time delay estimation errors have Gaussian distributions then the maximum likelihood estimate of the target position in the overlap region R_2 is given by

$$\hat{O}_{x'} = \frac{\sigma^2_{x'_2} R_{1_0} c\hat{\tau}_{12}/D_1 - \sigma^2_{x'_1}(c\hat{\tau}_{21}/2 - R_{2_0})}{\sigma^2_{x'_1}\sigma^2_{x'_2}} \tag{8.19}$$

$$\hat{O}_{y'} = \frac{\sigma^2_{y'_2}(c\hat{\tau}_{11}/2 - R_{1_0}) + \sigma^2_{y'_1} R_{2_0} c\tau_{22}/D_1}{\sigma^2_{y'_1}\sigma^2_{y'_2}} \tag{8.20}$$

From Figure 8.1 it is seen that the target position (O_x, O_y) can be found from the time delays at either sensor. The position coordinates are determined from

$$\begin{aligned} \hat{O}_x &= \hat{R}_1 \cos(\hat{\phi}_1) \\ &= R_{12}\cos(\alpha) + \hat{R}_2 \cos(\hat{\phi}_2) \end{aligned} \tag{8.21}$$

$$\begin{aligned} \hat{O}_y &= \hat{R}_1 \sin(\hat{\phi}_1) \\ &= R_{12}\sin(\alpha) + \hat{R}_2 \sin(\hat{\phi}_2) \end{aligned} \tag{8.22}$$

where $\hat{R}_i$ and $\hat{\phi}_i$ are obtained from (8.4, 8.5). Define the position error variances

$$\begin{aligned} \sigma^2_{x_1} &= \mathrm{Var}[\hat{R}_1 \cos(\hat{\phi}_1)] \\ \sigma^2_{x_2} &= \mathrm{Var}[R_{12}\cos(\alpha) + \hat{R}_2 \cos(\hat{\phi}_2)] \\ \sigma^2_{y_1} &= \mathrm{Var}[\hat{R}_1 \sin(\hat{\phi}_1)] \\ \sigma^2_{y_2} &= \mathrm{Var}[R_{12}\sin(\alpha) + \hat{R}_2 \sin(\hat{\phi}_2)] \end{aligned} \tag{8.23}$$

For small angles $\Delta\phi_i$ these error variances can be expressed in terms of the time delay variances through the use of (8.5) and (8.4). This yields

$$\begin{aligned} \sigma^2_{x_1} &= R^2_{1_0}\sin^2(\phi_{1_0})(c/D_1)^2\mathrm{Var}[\tau_{12}] + (c/2)^2\cos^2\phi_{1_0}\mathrm{Var}[\tau_{11}] \\ \sigma^2_{x_2} &= R^2_{2_0}\sin^2(\phi_{2_0})(c/D_1)^2\mathrm{Var}[\tau_{22}] + (c/2)^2\cos^2\phi_{2_0}\mathrm{Var}[\tau_{21}] \\ \sigma^2_{y_1} &= R^2_{1_0}\cos^2(\phi_{1_0})(c/D_1)^2\mathrm{Var}[\tau_{12}] + (c/2)^2\sin^2\phi_{1_0}\mathrm{Var}[\tau_{11}] \\ \sigma^2_{y_2} &= R^2_{2_0}\cos^2(\phi_{2_0})(c/D_1)^2\mathrm{Var}[\tau_{22}] + (c/2)^2\sin^2\phi_{2_0}\mathrm{Var}[\tau_{21}] \end{aligned} \tag{8.24}$$

In the overlap region the estimates can be combined to form the weighted estimate

$$\hat{O}_x = \frac{\sigma^2_{x_2}(\hat{R}_1\cos(\hat{\phi}_1)) + \sigma^2_{x_2}(R_{12}\cos(\alpha) + \hat{R}_2\cos(\hat{\phi}_2))}{\sigma^2_{x_1}\sigma^2_{x_2}} \tag{8.25}$$

$$\hat{O}_y = \frac{\sigma^2_{y_1}(\hat{R}_1\sin(\hat{\phi}_1)) + \sigma^2_{y_2}(R_{12}\sin(\alpha) + \hat{R}_2\sin(\hat{\phi}_2))}{\sigma^2_{y_1}\sigma^2_{y_2}} \tag{8.26}$$

The Doppler velocity estimate $\hat{\mathbf{V}}$ and Doppler angle estimates $\hat{\gamma}_1$ and $\hat{\gamma}_2$ can also be estimated in the overlap region. With the estimates $\hat{\phi}_1$ and $\hat{\phi}_2$ in hand, the

angle $\hat{\zeta}$ is found from

$$\hat{\zeta} = \hat{\phi}_2 - \hat{\phi}_1$$

Using the Doppler velocity equation

$$\hat{V}_{d_i} = |\hat{\mathbf{V}}|\cos(\hat{\gamma}_i), \tag{8.27}$$

and $\hat{\gamma}_1 = \hat{\gamma}_2 + \hat{\zeta}$, the ratio of Doppler velocities gives

$$\frac{\hat{V}_{d_1}}{\hat{V}_{d_2}} = \frac{\cos(\hat{\gamma}_2 + \hat{\zeta})}{\cos(\hat{\gamma}_2)}$$

Solving for $\hat{\gamma}_2$

$$\hat{\gamma}_2 = \tan^{-1}\left[\frac{\cos(\hat{\zeta}) - \hat{V}_{d_1}/\hat{V}_{d_2}}{\sin(\hat{\zeta})}\right]$$

With the estimate $\hat{\gamma}_2$, the magnitude of the Doppler velocity $|\hat{\mathbf{V}}|$ can be found from (8.27), and the target heading is $\hat{\psi} = -\hat{\gamma}_2 + \hat{\phi}_2 - \pi$.

8.2 Joint Detection/Estimation

The target search region has been localized to the rectangular box shown in Figure 8.1. This box is subdivided into several resolution cells as shown in this figure. The beam pattern from sensor s_1 allows this sensor to detect a target and estimate its parameters if the target is located in resolution cells 1 through 21. Sensor s_2 can detect the target if it is in cells 11 through 15, 22 through 25, or 26 through 31. If the target is not located in any of these cells then the target is declared not present (or more precisely, not detectable) . This situation is represented by the null hypothesis H_0. The resolution cells are grouped into regions which will be used for minimum mean square error estimation. If the target is located in regions R_1 (resolution cells 1 through 9) or R_3 (resolution cells 16 through 21) only

sensor s_1 can detect the target. Regions R_4 (resolution cells 22 through 25) and R_5 (resolution cells 26 through 31) correspond to the coverage area of sensor s_2 only. If the target is located in region R_2 (resolution cells 10 through 15) both sensors can detect the target and perform parameter estimation. In this case the estimates can be combined as described in section 8.1.2. The remaining area in the rectangle in Figure 8.1 is designated as region R_0, where neither sensor can detect the target.

The joint detection/estimation (JD/E) procedure applied to this problem involves both model uncertainty and uncertain initial conditions. JD/E for this situation is discussed in Chapter 5. Let $\theta_i \in \Theta$ designate the parameter vector that describes the different combination of model uncertainty and initial condition uncertainty. The parameter vector θ_i is assumed to be time invariant. The parameter vector θ_j, $j = (n-1) * M + m$, $1 \leq n \leq N$, $1 \leq m \leq M$, is defined to be the $(n,m)^{th}$ delay/Doppler resolution cell and is used to define $NM + 1$ different combinations initial conditions and models. n corresponds to the range resolution cell number determined from the initial conditions on the two time delays from each sensor, and m corresponds to the velocity resolution cell number determined from the initial conditions on the Doppler shift from each sensor. Since the model uncertainty is associated with the spatial coverage of each sensor, Doppler shift estimation is not considered in the experimental evaluation. The signal carrier frequency ω_c and pulse width t_w will be assigned values that lead to good delay estimation, but poor Doppler estimation. This is done so that large Doppler uncertainty has limited effect on the model used in each spatial resolution cell, and the focus can be directed primarily on range and azimuth resolution.

The number of Doppler resolution cells is set to $M = 1$ in order to simplify the discussion to follow. For the case $M > 1$ the initial conditions on Doppler shift

would be different for each hypothesis. This situation is discussed in Chapter 7. For $M = 1$ the parameter θ_j is directly associated with the range resolution cell number. Hypothesis H_{θ_j}, redesignated H_j, corresponds to the hypothesis that the target is present in range resolution cell j.

In region R_0, from which neither sensor can detect the target, hypothesis H_0 is defined by

$$H_0 \; : \; \mathbf{z}_{i_k} = \mathbf{u}_{i_k} + \mathbf{v}_{i_k} \qquad \forall k, \;\; i = 1,2 \tag{8.28}$$

For those resolution cells in regions R_1 and R_3, the hypothesis corresponding to cell $j, j = 1,2,\cdots,9$ (Region R_1), $j = 16,17,\cdots,21$ (Region R_3) is given by

$$H_j : \begin{cases} z_{1m_k} = \begin{cases} u_{1m_k} + v_{1m_k} & k\,t_s < \hat{\tau}_{1m_{j_k}} \\ h_{1m_k} + u_{1m_k} + v_{1m_k} & \hat{\tau}_{1m_{j_k}} \le k\,t_s < \hat{\tau}_{1m_{j_k}} + t_{w_1} \\ u_{1m_k} + v_{1m_k} & k\,t_s \ge \hat{\tau}_{1m_{j_k}} + t_{w_1} \end{cases} \\ \\ \mathbf{z}_{2_k} = \mathbf{u}_{2_k} + \mathbf{v}_{2_k} \qquad \forall k \end{cases} \tag{8.29}$$

for $m = 1,2$. The delay $\hat{\tau}_{i_{j_k}}$ is given by

$$\hat{\tau}_{im_{j_k}} = \hat{\tau}_{i1_{j_k}} + \kappa_m \hat{\tau}_{i2_{j_k}} \tag{8.30}$$

where $\kappa_m = +1$ whenever $m = 1$, and $\kappa_m = -1$ whenever $m = 2$. The initial conditions are given by

$$\begin{aligned} \hat{x}_{1_{0|0,\theta_j}} &= [\hat{\tau}_{11_{j_0}}, \hat{\tau}_{12_{j_0}}, \hat{\nu}_{1_0}]^T \\ P_{1_{0|0,\theta_j}} &= \text{Diag}\left[\text{Var}[\hat{\tau}_{11_{j_0}}], \text{Var}[\hat{\tau}_{12_{j_0}}], \text{Var}[\hat{\nu}_{1_0}]\right] \end{aligned} \tag{8.31}$$

The initial estimates $\hat{\tau}_{i1_{j_0}}, \hat{\tau}_{i2_{j_0}}, i = 1,2$ are chosen such that the position of the target for a signal received ate sensor i is at the center of resolution cell j. The

variances $\mathrm{Var}[\hat{\tau}_{i1_{j_0}}]$ and $\mathrm{Var}[\hat{\tau}_{i2_{j_0}}]$ are determined based on a uniform distribution of the error within the cell.

In region R_2 both signals are assumed to be present. In this region the hypothesis associated with cell j, $j = 12, 13, \cdots, 15$ is given by

$$H_j : \begin{cases} z_{1m_k} = \begin{cases} u_{1m_k} + v_{1m_k} & k\,t_s < \hat{\tau}_{1m_{j_k}} \\ h_{1m_k} + u_{1m_k} + v_{1m_k} & \hat{\tau}_{1m_{j_k}} \le k\,t_s < \hat{\tau}_{1m_{j_k}} + t_{w_1} \\ u_{1m_k} + v_{1m_k} & k\,t_s \ge \hat{\tau}_{1m_{j_k}} + t_{w_1} \end{cases} \\ \\ z_{2m_k} = \begin{cases} u_{2m_k} + v_{2m_k} & k\,t_s < \hat{\tau}_{2m_{j_k}} \\ h_{2m_k} + u_{2m_k} + v_{2m_k} & \hat{\tau}_{2m_{j_k}} \le k\,t_s < \hat{\tau}_{2m_{j_k}} + t_{w_2} \\ u_{2m_k} + v_{2m_k} & k\,t_s \ge \hat{\tau}_{2m_{j_k}} + t_{w_2} \end{cases} \end{cases} \tag{8.32}$$

for $m = 1, 2$. The initial conditions are given by

$$\begin{aligned} \hat{x}_{i_{0|0,\theta_j}} &= [\hat{\tau}_{i1_{j_0}}, \hat{\tau}_{i2_{j_0}}, \hat{\nu}_{i_0}]^T \\ P_{i_{0|0,\theta_j}} &= \mathrm{Diag}\left[\mathrm{Var}[\hat{\tau}_{i1_{j_0}}], \mathrm{Var}[\hat{\tau}_{i2_{j_0}}], \mathrm{Var}[\hat{\nu}_{i_0}]\right] \\ & \qquad\qquad i = 1, 2 \end{aligned} \tag{8.33}$$

For those resolution cells in regions R_4 and R_5, where only sensor 2 can detect the target, the hypothesis corresponding to cell j, $j = 22, 23, \cdots, 25$(Region R_4), $j = 26, 27, \cdots, 31$(Region R_5) is given by

$$H_j : \begin{cases} \mathbf{z}_{1_k} = \mathbf{u}_{1_k} + \mathbf{v}_{1_k} \quad \forall\, k \\ \\ z_{2m_k} = \begin{cases} u_{2m_k} + v_{2m_k} & k\,t_s < \hat{\tau}_{2m_{j_k}} \\ h_{2m_k} + u_{2m_k} + v_{2m_k} & \hat{\tau}_{2m_{j_k}} \le k\,t_s < \hat{\tau}_{2m_{j_k}} + t_{w_2} \\ u_{2m_k} + v_{2m_k} & k\,t_s \ge \hat{\tau}_{2m_{j_k}} + t_{w_2} \end{cases} \end{cases} \tag{8.34}$$

for $m = 1, 2$. The initial conditions are given by

$$\begin{aligned} \hat{x}_{2_{0|0,\theta_j}} &= [\hat{\tau}_{21_{j_0}}, \hat{\tau}_{22_{j_0}}, \hat{\nu}_{2_0}]^T \\ P_{2_{0|0,\theta_j}} &= \text{Diag}\left[\text{Var}[\hat{\tau}_{21_{j_0}}], \text{Var}[\hat{\tau}_{22_{j_0}}], \text{Var}[\hat{\nu}_{2_0}]\right] \end{aligned} \tag{8.35}$$

A maximum a posteriori detection criterion can be used to determine the most likely range resolution cell. This criterion requires the availability of *apriori* probabilities of each hypothesis, and it requires the probability density functions for the measurements. Define $\mathbf{Z}_k = [\mathbf{z}_1, \mathbf{z}_2, \cdots \mathbf{z}_k]$, where $\mathbf{z}_k = [\mathbf{z}_{1_k}^T, \mathbf{z}_{2_k}^T]^T$, as the set of all measurements up to time k, and let $p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_j)$ be the probability density function of $\mathbf{z}_k$ given the measurements $\mathbf{Z}_{k-1}$ and hypothesis H_j. The a *posteriori* probability of hypothesis H_i is given by

$$P(\theta_j|\mathbf{Z}_k) = \frac{P(\theta_j|\mathbf{Z}_{k-1})\,\Lambda_i(\mathbf{z}_k)}{\sum_{m=0}^{N} P(\theta_m|\mathbf{Z}_{k-1})\,\Lambda_m(\mathbf{z}_k)} \tag{8.36}$$

where $\Lambda_j(\mathbf{z}_k)$ is the likelihood ratio defined by

$$\Lambda_k(\mathbf{z}_k) = \frac{p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_k)}{p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_0)} \tag{8.37}$$

In general the distribution function $p(\mathbf{z}_k|\mathbf{Z}_{k-1}, \theta_i)$ is non-Gaussian. Since the measurement noise consists of a sum of Gaussian noise and non-Gaussian clutter, the joint density function consists of a convolution of the Gaussian and non-Gaussian density functions. In general it is not possible to compute this joint density analytically and must be done numerically for each iteration of the filter, since the density function changes as the estimate $\hat{\mathbf{x}}_{k|k,\theta_i}$ changes.

Maximum a *posteriori* (MAP) detection can be used to decide the most likely

hypothesis according to :

$$\text{Choose} \quad H_j \quad : \quad \theta_j = \text{argmax}_{\theta_m \in \Theta} P(\theta_m | \mathbf{Z}_k) \tag{8.38}$$

The MAP estimate from sensor s_i is then the estimate associated with cell j if cell j is in the spatial area of coverage of that sensor.

The minimum mean square error estimate can be found be combining the estimates from all of the cells with a particular region. If the state vector $\mathbf{x}_k$ is common to all models the minimum mean squared error (MMSE) estimate can be used. The MMSE estimate for sensor i in region R_p can be expressed by

$$\hat{\mathbf{x}}^*_{i_{k|k}} = \sum_{\text{cell}_j \in R_p} P(\theta_j | \mathbf{Z}_k) \hat{\mathbf{x}}_{i_{k|k,\theta_j}}. \tag{8.39}$$

The most likely region is selected using the MAP criterion. Define as the hypothesis that the target is located in region R_p as I_p, $p = 0, 1, \cdots, 5$. The *a posteriori* probability associated with region R_p is the sum of the *a posteriori* probabilities of all of the cells in that region. This region-level probability is given by

$$P(I_p | \mathbf{Z}_k) = \sum_{\text{cell}_j \in R_p} P(\theta_j | \mathbf{Z}_k) \tag{8.40}$$

The most likely region is chosen such that

$$\text{Choose} \quad I_p \quad : \quad p = \text{argmax}_{p=0,\cdots,5, \theta_j \in \Theta} P(I_p | \mathbf{Z}_k) \tag{8.41}$$

8.2.1 Definition of Priors

The *a priori* probabilities of each hypothesis are based on the area coverage of the sensors. The total number of resolution cells shown in Figure 8.1 is 56. Of

these, 25 are located in region R_0. All cells are assumed to have an equal probability containing the target. The a *priori* probabilities are given by

$$\begin{aligned} P(\theta_0) &= 25/56 \\ P(\theta_j) &= 1/56, \qquad j = 1, 2, \cdots 31 \end{aligned} \tag{8.42}$$

The probabilities associated with regions R_j, $j = 0, 1, \cdots, 5$ are given by

$$\begin{aligned} P(I_0) &= 25/56 \\ P(I_1) &= 9/56 \\ P(I_2) &= 6/56 \\ P(I_3) &= 6/56 \\ P(I_4) &= 4/56 \\ P(I_5) &= 6/56 \end{aligned} \tag{8.43}$$

8.3 Simulation Experiments

An experimental study was conducted to evaluate the performance of the multisensor fusion technique. In this evaluation the measurement noise consisted of 50% Lognormal Noise and 50% Gaussian noise. The nominal angles from sensors s_1 and s_2 to the target were $\phi_{1_0} = 45\,\text{deg}$ and $\phi_{2_0} = 135\,\text{deg}$, respectively. The nominal range from s_1 to the target was 10 miles. The nominal range from sensor s_2 to the target was chosen such that the received signal at s_2 was 5 dB higher than at s_1 for the same transmitted signal level and target strength.

The carrier frequencies used by the two sensors were the same at $f_c = 10 \times 10^6$. This is not practical situation since two cooperating sensors would not transmit at

the same frequency, unless they use the same signal generator and transmissions from the two sensors are offset in time. However, it is desired to show the effect of only one variable, the relative SNR at each sensor, on the estimation error. Since the operating frequency affects the variance of the estimates as described by (7.31), the operating frequencies are kept the same. Both sensors illuminate the target simultaneously. They both sample the signal at a rate $f_s = 100 \times 10^6$, and both signals have the same pulse width $t_{w_i} = 12/f_s$, $i = 1, 2$. The resolution cell width is $1/f_s$ seconds. The associated initial error variance on time delays τ_{11_0} and τ_{21_0} is $t_s^2/12$. The corresponding range resolution cell width is $\Delta r_i = c/(2f_s)$. Thus, the initial variance for the angle-measurement delays is (8.6) $\text{Var}[\tau_{12_0}] = ((D_i c)/(2f_s R_i))^2/12$, $i = 1, 2$. D_i, the separation between phase centers at the sensor was chosen to be 3 feet for each sensor.

The carrier frequencies, pulse widths, and sampling frequencies chosen for this evaluation are the same as that chosen for the time delay estimation experiment in Chapter 7. It was observed in Chapter 7 that the values chosen for these parameters are not conducive to estimation of Doppler shift. Since the primary goal of the evaluation in this chapter is to properly locate the correct region and resolution cell number, the estimation of Doppler shift plays a secondary role. Accordingly, Doppler shift is not estimated in this evaluation. Target positions are selected randomly with a uniform distribution in the spatial area designated by the large box in Figure 8.1.

Simulations were performed for SNR's (at sensor s_1) ranging from -10dB to 10dB. 500 random target positions were chosen at each SNR. Of these 500 trials, 228 target positions randomly chosen in region R_0, 91 in R_1, 54 in R_2, 44 in R_3, 40 in R_4, and 40 in R_5.

Table 8.1 shows the detection results for the EKF using the Gaussian pdf to evaluate the *a posteriori* density function in (8.37). The average *a posteriori* probability for correct decisions at the region level is given by $\hat{P}(I_p|\mathbf{Z}_k)$. This is computed as the arithmetic mean of the *a posteriori* probabilities (8.40) for those trials in which the correct region was chosen using (8.41). This value gives may be used as a measure of the level of confidence that the proper region was chosen. At the resolution cell level the average probability is denoted $\hat{P}(\theta_j|\mathbf{Z}_k)$. A target was declared present if the *a posteriori* probability $P(I_p|\mathbf{Z}_k)$ for any region $p, p = 1, \cdots, 5$ was greater than $P(I_0|\mathbf{Z}_k)$. The probability of detection is labeled $P(I_p|I_q)$, $p, q \neq 0$. This quantity was determined by dividing the total number of declared detections, or the number of trials in which a target was declared present in any of the regions R_1 through R_5, by the total number of trials in which the target was actually located in one of the regions R_1 through R_5. The probability of false alarm, $P(I_p|I_0)$, was determined by dividing the total number of trials in which a target was declared present when it was actually in R_0, by the number of trials in which the target was actually in R_0.

Table 8.1. Multisensor Fusion Detection and False Alarm Probabilities

SNR(dB)	$\hat{P}(\theta_j\|\mathbf{Z}_k)$	$\hat{P}(I_p\|\mathbf{Z}_k)$	$P(I_p\|I_q)$, $p, q \neq 0$	$P(I_p\|I_0)$
-10	0.61	0.74	0.74	0.18
-5	0.77	0.93	0.91	0.059
0	0.83	1.0	0.99	0.0044
5	0.84	1.0	1.0	0.0
10	0.84	1.0	1.0	0.0

The JD/E technique performs very well in terms of locating the proper region. However, the performance is not as good in finding the correct cell. This is due to the fact that the initial variance in angle delay is so small that the filter cannot properly decide the correct cell number. The probabilities of missed detection $P(I_0|I_p)$ and correct classification (i.e. not only detection of the target but correct localization at the region level) $P(I_p|I_p)$, $p = 1, \cdots, 5$ are displayed in Table 8.2. The probability of misclassification, which is not shown in this table, is given by $P(I_q|I_p) = 1 - P(I_p|I_p) - P(I_0|I_p)$, $q \neq p$. Sensor s_2 outperforms sensor s_1, which is to be expected since the SNR at s_1 is 5 dB higher than the SNR at sensor s_2. In the overlap region, R_2, the classification performance is much better than it is for any other region, with an 85% probability of correct classification.

Table 8.2. Probabilities of Missed Detection and Correct Classification - Region Level

SNR(dB)		Probability				
		$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
-10	$P(I_0\|I_p)$	0.35	0.074	0.50	0.15	0.16
	$P(I_p\|I_p)$	0.57	0.85	0.45	0.78	0.79
-5	$P(I_0\|I_p)$	0.13	0.019	0.23	0.025	0.023
	$P(I_p\|I_p)$	0.87	0.96	0.77	0.98	0.98
0	$P(I_0\|I_p)$	0.022	0.0	0.023	0.0	0.0
	$P(I_p\|I_p)$	0.98	1.0	0.98	1.0	1.0
5	$P(I_0\|I_p)$	0.0	0.0	0.0	0.0	0.0
	$P(I_p\|I_p)$	1.0	1.0	1.0	1.0	1.0
10	$P(I_0\|I_p)$	0.0	0.0	0.0	0.0	0.0
	$P(I_p\|I_p)$	1.0	1.0	1.0	1.0	1.0

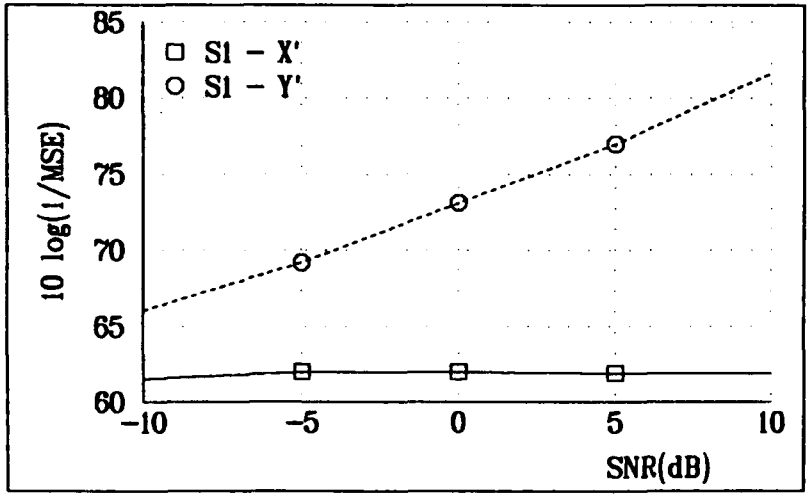
The missed detection and correct classification probabilities at the cell level are shown in Table 8.3. The results are averaged over all of the cells in each region. Again the performance for those cells in region R_2 was much better than for any other region. The classification results for regions R_1, R_3, R_4, and R_5 were poor even at high SNR's. The results in this table reflect the inability of the sensors to detect the proper cell number in the cross-range direction. Figure 8.1 shows that there are three cells in the cross range direction for sensor s_1, and two cells for sensor s_2. Thus assuming that the cell number cannot be resolved in the cross-range direction, the expected cross-range uncertainty for regions R_1 and R_3 is 1/3, and the cross-range uncertainty for R_4 and R_5 is 1/2. This is verified by the experimental results in Table 8.3 at 10 dB SNR.

Table 8.3. Probabilities of Missed Detection and Correct Classification - Cell Level

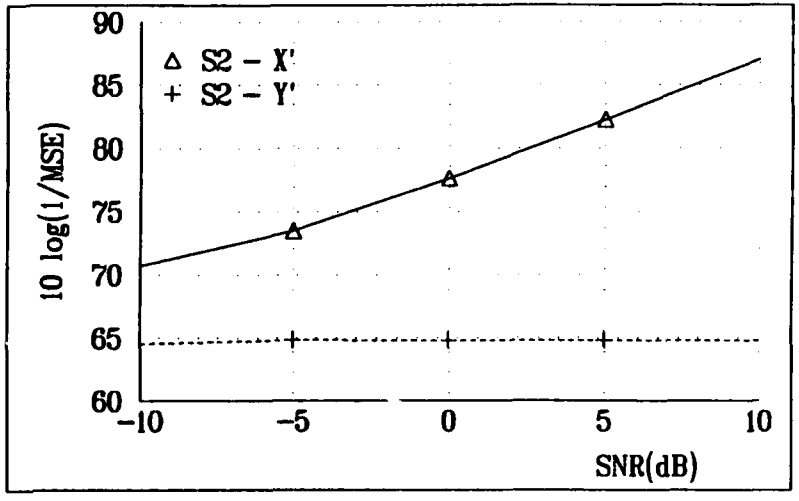
SNR(dB)		Probability				
		R_1	R_2	R_3	R_4	R_5
-10	$P(\theta_0\|\theta_j)$	0.85	0.20	0.86	0.28	0.33
	$P(\theta_j\|\theta_j)$	0.022	0.52	0.046	0.23	0.21
-5	$P(\theta_0\|\theta_j)$	0.29	0.019	0.39	0.025	0.023
	$P(\theta_j\|\theta_j)$	0.16	0.76	0.20	0.33	0.44
0	$P(\theta_0\|\theta_j)$	0.033	0.0	0.046	0.0	0.0
	$P(\theta_j\|\theta_j)$	0.33	0.94	0.27	0.43	0.51
5	$P(\theta_0\|\theta_j)$	0.0	0.0	0.0	0.0	0.0
	$P(\theta_j\|\theta_j)$	0.39	0.94	0.27	0.45	0.53
10	$P(\theta_0\|\theta_j)$	0.0	0.0	0.0	0.0	0.0
	$P(\theta_j\|\theta_j)$	0.36	0.96	0.32	0.45	0.56

The estimation results are shown in Figure 8.3. All results shown in this figure are in reference to the (X', Y') coordinate system. Figure 8.3(a) shows the average mean squared error for those detections in regions R_1 and R_3, in which only s_1 has coverage. The results in this figure are consistent with those in Table 8.3 in that the estimates in the cross-track direction X' never improve over the initial estimates regardless of the SNR. Figure 8.3(c) shows similar results for regions R_4 and R_5, which are covered by sensor s_2. Figure 8.3(c) also illustrates the 5 dB performance for sensor s_2 over that for s_1. Figure 8.3(b) shows the results for both sensors in region R_2. In this region, as shown in Table 8.3 the proper cell is almost always found. Thus the cross-range estimation error variance should improve by about 6 dB (20log(2)) for sensor s_2, since the cross-range error for s_2 has been localized from 2 cells down to 1. Similarly, the cross-range error variance for sensor s_1 in Region R_2 is reduced by about 10 dB (20log(3)) since the target has been localized from 3 cells down to 1. This improvement is evident in Figure 8.3(b). Figure 8.3(d) shows the estimation results using the combined measurents obtained from (8.19and 8.20). Because of the larger variance in the cross-range error for each sensor and the fact that the intersection of the LOS's between the two sensors are perpendicular, the combined estimate consists of the X' estimate from sensor s_2 and the Y' estimate from sensor s_1.

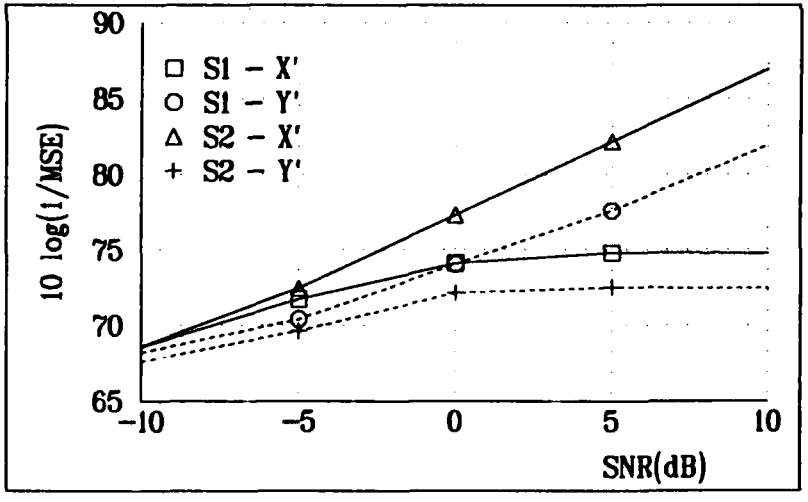
The mean squared errors in the (X, Y) coordinate system are shown in Figure 8.4. Each curve in this figure represents the combined X and Y position errors, since the geometry dictates that the error variance should be the same in each direction. The (X, Y) positions are obtained using (8.21) and (8.22). This figure again illustrates the approximate 5 dB improvement in the estimates from sensor s_2 over that of sensor s_1 in the nonoverlapped regions, and the significant performance



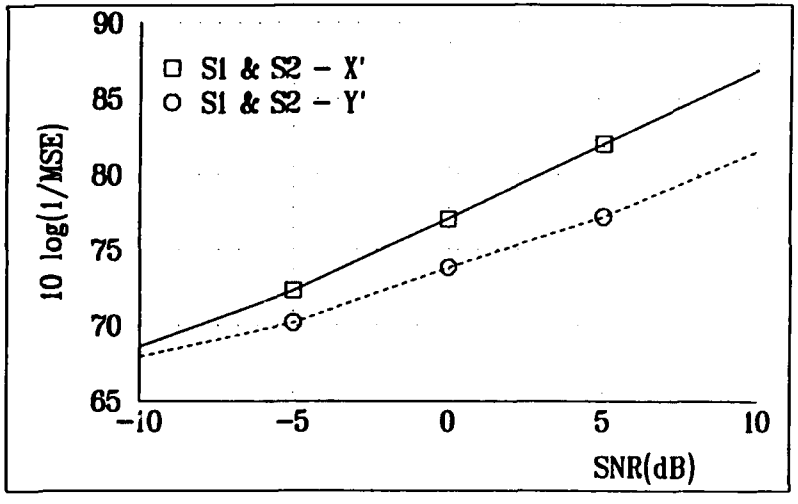
(a) Errors from s_1 in Regions R_1 & R_3



(c) Errors from s_2 in Regions R_4 & R_5



(b) Errors from s_1 and s_2 in Region R_2



(d) Combined Errors from s_1 and s_2 in Region R_2

Figure 8.3 Multisensor Fusion In-Track and Cross-Track Estimation Errors

improvement in region R_2. It is observed that the (X, Y) position errors for the overlap region shown in Figure 8.4 are significantly worse than those shown in Figure 8.3(d) for the (X', Y') coordinate system. This is particularly evident at high SNR's. This is due to the fact that the cross-track errors are included in the computation of the combined estimate O_x and O_y determined by (8.21) and (8.22). The choice of the proper coordinate system can make a large impact on the performance of the estimator.

8.4 Conclusion

A technique has been presented for multisensor fusion based on joint detection/estimation procedure. It is shown that excellent performance can be obtained for both target detection and target parameter estimation using this technique. A significant advantage of this technique is that each sensor can perform detection and parameter estimation in a decentralized mode. The final estimates and *a posteriori* probabilities from each sensor are processed by a centralized processor to derive the optimum estimate.

The method provides an automatic referencing mechanism of the data from the different sensors (automatic data alignment) as long as the geometry and timing of the sweeping beams are known. For optimal target resolution performance, it is found that the lines of sight of the two sensors should be perpendicular to each other at any given time, requiring special synchronization. This implies that if the sweeping angle of one of the sensors, e.q. s_1 as a function of time is $\phi_{1_0}(t)$, the corresponding sweeping angle of sensor s_2 must be $\phi_{2_0}(t) = \pi/2 + \phi_{1_0}(t)$, a goal that is easily accomplished with an efficient model reference (adaptive) controller.

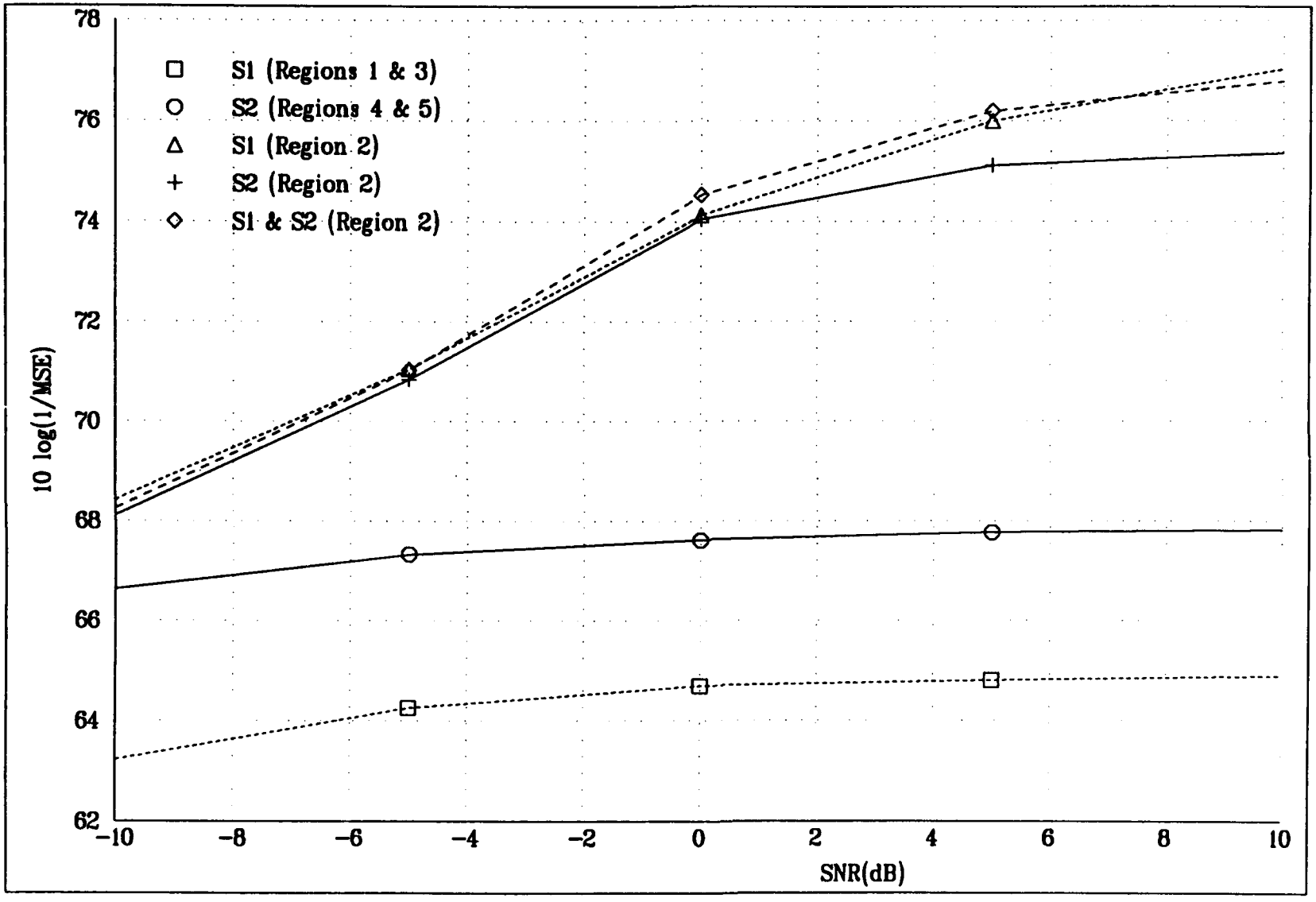


Figure 8.4 X and Y Position Errors for Single and Multiple Sensors

Chapter 9

Summary and Areas for Further Study

Two high order filters (HOFs) have been presented for estimation in non-Gaussian noise. The first filter is designed for systems with asymmetric probability densities. The asymmetrical filter is developed by using the first and second powers of the innovations in the derivation of the filter equations. The second filter is designed for systems with symmetric probability densities. It is developed based on first and third powers of the innovations. These filters are evaluated experimentally in non-Gaussian noise formed from Gaussian sum distributions. Under these conditions the HOFs perform much better than the standard Kalman filter, and close to the optimal Bayesian estimator, the Gaussian sum filter. However, the primary advantage of using the HOFs occurs either when the noise cannot be adequately represented as Gaussian sums, or when only the moments of the noise are known, and not the actual density functions. Although these filters are more complicated to implement than the standard Kalman filter, they are not nearly as computationally intensive as the Gaussian sum filter for which the number of parallel filters grows geometrically as the number of stages increase.

For HOFs designed for I^{th} order filter moments, their implementation requires the availability of prediction error moments of order up to $2I$. In general, when $I > 1$ it is necessary to either truncate the expressions for the filter moments so that only those powers of prediction and measurement error moments are included for which similar powers of the filter moments exist, or the higher powers of the prediction and measurement error moments must be approximated. This leads to either the truncation of the filter error moment expressions or the estimation of

prediction error moments of order $I+1$ through $2I$. It is shown that the truncated filter expressions give comparable performance to those with estimated higher order moments.

For non-Gaussian distributions made up of known Gaussian sums, the non-Gaussian filters presented here give a reasonable compromise between the optimal but very computationally intensive Gaussian sum filter, and the suboptimal but easily implemented standard Kalman filter. In addition, when only the moments of the distributions are known and a Gaussian sum filter cannot be used, the non-Gaussian filters offer a means to obtain improved performance over the standard Kalman filter. One method to improve the performance of the non-Gaussian filters is to use higher powers of the innovations in developing the filter equations. However, the resulting filter expressions would be extremely complicated and it is anticipated that the expected performance improvement over the HOFs presented here may be marginal.

A more general filter can be developed by including the first, second, and third order powers of the innovations in developing the filter equations. This can be useful, for example, in a situation in which the measurement noise has an asymmetrical distribution and the process noise has a symmetrical non-Gaussian distribution. The derivation of this filter will follow the same procedure as shown in Chapter 3. Three separate gain matrices will be required in this case.

From the implementation standpoint a significant reduction in the computational burden imposed by the HOFs can be accomplished by exploiting the redundancy in the high order filter moment matrices. For example the error covariance matrix can be represented by either the upper or lower triangular matrix. Similarly,

matrices for the 3^{rd} and 4^{th} order expansions also contain a significant amount of redundancy, and efficient algorithms may be developed for including only the necessary terms in these matrices.

This thesis also addresses several signal processing estimation problems with a model-based formalism. These problems are all treated as nonlinear estimation problems in Gaussian and non-Gaussian noise. A direct model is used in which the frequencies, amplitudes, damping coefficients and phases of the sinusoids are defined as state variables. This model has the advantage that the time varying behavior of these parameters can be directly described through the process equation. The harmonic retrieval problem is solved using three separate nonlinear filters and three iterated forms of the extended Kalman filter. The nonlinear filters offer a significant advantage over batch-type estimators in that time varying system parameters can be modeled. A problem that has been studied by other authors is addressed and it is found that the nonlinear filters offer a significant advantage over other techniques such as modified singular value decomposition and cumulant-based techniques whenever the initial estimation error is constrained. It is shown that the nonlinear filters can be used effectively in colored Gaussian noise with known or unknown coefficients, and in measurement noise with known and unknown covariances. Another advantage of the nonlinear filter approach is that these filters converge relatively fast, making them well-suited for short data lengths.

A joint detection/estimation (JD/E) procedure is presented and applied to problems with model uncertainty and/or uncertain initial conditions. The implementation of this procedure consists of several filters operating in parallel. Each filter hypothesizes a different measurement or process model, different initial conditions, or both. The estimators act independently of the detection mechanism. The

link between the two is provided by the *a posteriori* probability, which is evaluated for any arbitrary density function. These estimators can include any recursive filter such as the linear Kalman filter, nonlinear filters, or the HOFs.

The JD/E approach is applied to model order selection. A general approach is presented for determining the number of sinusoids present in measurements corrupted by additive white Gaussian and non-Gaussian noise. The approach involves the simultaneous application of *maximum a posteriori* (MAP) detection and nonlinear estimation of the state variables, which consist of the amplitudes and frequencies of sinusoids in each model. Estimation is performed using the extended Kalman filter when the noise is Gaussian, and the extended high order filter (EHOF) when the noise is in non-Gaussian. The initial state estimates are constrained to be within an initial variance. The problem is formulated as a multiple hypothesis testing problem with assumed known *a priori* probabilities for each hypothesis. Each hypothesis represents a different model. Experimental evaluation of this approach demonstrates excellent performance for model order selection and system parameter estimation in both Gaussian and non-Gaussian noise.

The JD/E approach for problems with uncertain initial conditions is applied to the estimation of the time delay and Doppler shift from the active echo returns of monostatic sensor(s). The problem becomes one of localizing a target in range-Doppler space. The range-Doppler space is partitioned into a number of resolution cells. Each cell is identified with a hypothesis that the signal is present in it. The joint detection/estimation scheme is then used to localize the target and refine its parameter estimates (i.e. time delay and Doppler shift). The measurements that are used to localize the target consist of signal returns corrupted by additive white Gaussian and non-Gaussian noise. It is found that the initial estimation error for

time delay must be within $1/2f_c$ for any given estimator to form an accurate estimate of target position. This requirement can be relaxed with the JD/E scheme, since a very large initial estimation error can be segmented into a number of filters, each with a much smaller error. The MAP estimate gives very good results under these conditions.

The JD/E approach for the combination of model uncertainty and uncertain initial conditions is applied to the problem of data fusion from two cooperating, non-collocated sensors that are attempting to detect a target and estimate its position. The geometric areas of coverage of the two sensors partially overlap. Thus, the model is general enough to include sensor misalignment. In the overlap region the estimates from the two sensors are combined to produce improved estimates over the single sensor estimates.

Several hypotheses are postulated for detection. Each hypothesis corresponds to the ability of each sensor to detect the target in its area of coverage. The a *priori* probabilities of each decision is based on the area of coverage of the two sensors. For each hypothesis, a high order filter recursively estimates time delay, Doppler shift and geometric angle to the target from processing the returns of the transmitted signal from each sensor. These estimates are in turn used to estimate target position and velocity. For each of these hypotheses, another set of parallel filters is used to obtain more accurate estimates of signal parameters and to account for the stability problems that result from the first order Taylor series expansion used in the nonlinear filtering algorithms. This is accomplished by operating a separate filter for each of several different initial time delay estimates of the return signal. The maximum likelihood estimate for a given hypothesis is then determined as a weighted sum of the estimates from each of the local hypotheses, with the a

posteriori probability being used as the weighting function. It is assumed that the signals are imbedded in Gaussian noise and clutter. The clutter is treated as non-Gaussian noise with a lognormal or Weibull distribution. Excellent performance is obtained using the JD/E approach with high detection probability and very good target position estimates.

The restriction of small initial estimation error, made for the harmonic retrieval and model order selection problems can be relaxed, if the JD/E approach is used for estimation (the model with uncertain initial conditions). The model order selection initial conditions can also be relaxed if the JD/E approach for model uncertainty and uncertain initial conditions is used.

Since the estimation for each of the hypotheses in the JD/E approach is performed independently, this scheme is a natural application for parallel processing. The model selection or detection decision can be made by a centralized processor after all of the data is processed. Thus, the JD/E approach is very well suited for real-time implementation using advanced massively parallel computer architectures.

References

1 G. N. Sardis, *Self-Organizing Control of Stochastic Systems,* Marcel Dekker, Inc., New York, NY, 1977.

2 Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association,* Analytic Press, Inc., 1988.

3 S. C. Kramer and H. J. Sorenson, "Bayesian Parameter Estimation," *IEEE Trans. Automatic Contr.,* vol. AC-17, no. 4, pp. 439–448, August, 1972.

4 R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans. of the ASME, Journal of Basic Engineering,* pp. 35–45, March, 1960.

5 Y. C. Ho and R. C. K. Lee, "A Bayesian Approach to Problems in Stochastic Estimation and Control," *IEEE Trans. Automatic Contr.,* vol. AC-9, pp. 333–339, October, 1964.

6 A. Gelb (Ed), *Applied Optimal Estimation,* The Analytic Sciences Corporation, 1974.

7 T. R. Kronhamn, "Geometric Illustration of the Kalman Filter Gain and Covariance Update Algorithms," *IEEE Contr. Syst. Mag.,* pp. 41–43, May, 1985.

8 C. K. Chui and G. Chen, *Kalman Filtering with Real-Time Applications,* Springer-Verlag, Berlin, 1987.

9 A. H. Jazwinski, *Stochastic Processes and Filtering Theory,* Academic Press, New York, NY, 1970.

10 H. W. Sorenson and D. L. Alspach, "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica,* vol. 7, pp. 465–479, 1971.

11 D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," *IEEE Trans. Automatic Contr.,* vol. AC-17, no. 4, pp. 439–448, August, 1972.

12 H. W. Sorenson and A. R. Stubberud, "Non-linear Filtering by Approximation of the *a posteriori* Density," *International Journal of Control,* vol. 18, pp. 33–51, 1968.

13 R. S. Bucy and K. D. Senne, "Realization of Discrete-Time Nonlinear Estimators," *Proc of the Symp. on Nonlinear Estimation Theory and Its App.,* pp. 6–17, San Diego, CA, September 21-23, 1970.

14 R. J. P. Figueiredo and Y. G. Yan, "Spline Filters," *Proc of the 2nd Symp. on Nonlinear Estimation Theory and Its App.*, pp. 127–138, San Diego, CA, September 13-15, 1971.

15 M. L. Andrade, "Interpolation by Splines; Applications in Nonlinear Filtering of Dynamical Systems," Ph.D dissertation, Universidade Estadual de Campinas, Campinas, S. P. Brazil, 1975.

16 T. Kailath, "An Innovations Approach to Least-Squares Estimation. Part I: Linear Filtering in Additive White Noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, no.6, 646–655, December, 1968.

17 G. B. Di Masi, W. J. Runggaldier, and B. Barozzi, "Generalized Finite-Dimensional Filters in Discrete Time," *in Nonlinear Stochastic Problems,* (R. S. Bucy and J. M. F. Moura eds), pp. 267–277, D. Reidel Publishing Company, Dordrecht, Holland, 1983.

18 A. H. Jazwinski, "Filtering for Nonlinear Dynamic Systems," *IEEE Trans. Automat. Cont.*, vol. 11, pp. 765–766, 1966.

19 R. W. Bass, V. D. Norum, and L. Schwartz, "Optimal Multichannel Nonlinear Filtering," *Journal Math. Anal. App.*, vol. 16, pp. 152–166, 1966.

20 M. Athans, R. P. Wishner, and A. Bertolini, "Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements," *1968 Joint Automat. Cont. Conf.*, Ann Arbor, Michigan, pp. 364–382, June, 1968.

21 W. F. Denham and S. Pines, "Sequential Estimation when Measurement Function Nonlinearity is Comparable to Measurement Error," *AIAA Journal,* vol. 4, pp. 1071–1076, 1966.

22 M. L. Andrade Netto, L. Gimeno, and J. Mendes, "On the Optimal and Suboptimal Nonlinear Filtering Problem for Discrete-Time Systems," *IEEE Transactions on Automatic Control,* vol. AC-23, pp. 1062–1067, December, 1978.

23 D. F. Liang and G. S. Christensen, "New estimation Algorithms for Discrete Nonlinear Systems and Observations with Multiple Time Delays," *Int. Journal of Cont.*, vol. 23, no. 5, pp. 613–625, 1976.

24 D. F. Liang, "Exact and Approximate State Estimation Techniques for Nonlinear Dynamic Systems," in *Control and Dynamic Systems,* (C. T. Leondes, Ed.) vol. 19, pp. 2–80, Academic Press, New York, 1983.

25 D. F. Liang, "Comparisons of Nonlinear Recursive Filters for Systems with Nonnegligible Nonlinearities," in *Control and Dynamic Systems,* (C. T. Leondes, Ed.) vol. 20, pp. 341–401, Academic Press, New York, 1983.

26 C. J. Masreliez, "Approximate Non-Gaussian Filtering with Linear State and Observation Relations," *IEEE Trans. Automatic Contr.*, vol. AC-20, no. 1, pp. 107–110, February, 1975.

27 T. S. Rao and M. Yar, "Linear and Non-Linear Filters for Linear, but not Gaussian Processes," *Int. J. Control,* vol. 39, no. 1, pp. 235–246, 1984.

28 E. I. Verriest, "Linear Filters for Linear Systems with Multiplicative Noise and Nonlinear Filters for Linear System with Non-Gaussian Noise," *IEEE American Control Conf.*, vol. I, pp. 182–184, June, 1985.

29 F. E. Daum, "A New Nonlinear Filtering Formula for Non-Gaussian Discrete-Time Measurements," *Proc. 25th Conf. on Decision and Contr.*, vol. 2, pp. 1030–1031, December, 1986.

30 W. J. Vetter, "Matrix Calculus Operations and Taylor Expansions," *SIAM Review,* vol. 15, no. 2, pp. 352–369, April, 1973.

31 S. L. Marple, Jr., *Digital Spectral Analysis,* Prentice Hall, Inc., Englewood Cliffs, NJ, 1987.

32 J. Capon, "High Resolution Frequency-Wavenumber Spectrum Analysis," *Proc. IEEE,* vol. 57, no. 8, pp. 1408–1419, August, 1969.

33 C. Foias, A. E. Frazho, and P. J. Sherman, "A Geometric Approach to the Maximum Likelihood Spectral Estimator for Sinusoids in Noise," *IEEE Trans. Info. Theory,* vol. 34, no. 5, pp. 1066–1070, 1984.

34 S. M. Kay and S. L. Marple, "Spectrum Analysis – A Modern Perspective," *Proc. IEEE,* vol. 69, no. 11, pp. 1380–1418, November, 1981.

35 T. J. Ulrych and R. W. Clayton, "Time Series Modeling and Maximum Entropy," *Physics of Earth and Plan. Int.*, vol. 12, pp. 188–200, 1976.

36 J. A. Cadzow, "Autoregressive Moving Average Spectral Estimation: A Model Equation Error Procedure," *IEEE Trans. Geoscience and Remote Sensing,* vol. GE-19, no. 1, pp. 24–28, December, 1981.

37 S. W. Lang and J. H. McClellan' "Frequency Estimation with Maximum Entropy Spectral Estimators," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-28, no. 6, pp. 716–724, December, 1980.

38 T. K. Citron, T. Kailath, and V. U. Reddy, "A Comparison of Some Spectral Estimation Techniques for Short Data Lengths," *IEEE Int. Conf. Acoust. Speech, and Signal Processing*, vol. 3, pp. 2094–2097, 1982.

39 D. W. Tufts and R. Kumaresan, "Singular Value Decomposition and Improved Frequency Estimation Using Linear Prediction," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-30, no. 4, pp. 671–675, August, 1982.

40 D. W. Tufts and R. Kumaresan, "Estimation of Frequencies of Multiple Sinusoids Making Linear Prediction Perform Like Maximum Likelihood," *Proc. IEEE*, vol. 70, pp. –, September, 1982.

41 R. Kumaresan and D. W. Tufts, "Estimating the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-30, no. 6, pp. 833–840, December, 1982.

42 R. Kumaresan, "On the Zeros of the Linear prediction-Error Filter for Deterministic Signals," *IEEE Trans. Acoust., Speech, and Signal Porcessing*, vol. ASSP-31, no. 1, pp. 217–220, February, 1983.

43 R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 3, pp. 637–639, June, 1986.

44 Y. Hua and T. K. Sarkar, "Perturbation Analysis of TK Method for Harmonic Retrieval Problems," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 36, no. 2, pp. 228–240, February, 1988.

45 Y. Bresler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-34, pp. 1081–1089, October, 1986.

46 M. Li, "Adaptive Retrieval and Enhancement of Sinusoids in a Quadrature Sampling System Using a Decomposed Complex IIR Notch Filter," *Computers and Electrical Engineering*, vol. 18, no. 3/4, p.195, 1992.

47 T. K. Sarkar, D. K. Weiner, V. K. Jain, and S. A. Dianat, "Approximation by a Sum of Complex Exponentials Utilizing the Pencil of Function Method," *in Nonlinear Stochastic Problems*, (R. S. Bucy and J. M. F. Moura eds), pp. 87–100, D. Reidel Publishing Company, Dordrecht, Holland, 1983.

48 T. K. Sarkar, D. K. Weiner, V. K. Jain, and S. A. Dianat, "Approximation by a Sum of Complex Exponentials Utilizing the Pencil of Function Method," *in Nonlinear Stochastic Problems*, (R. S. Bucy and J. M. F. Moura eds), pp. 87–100, D. Reidel Publishing Company, Dordrecht, Holland, 1983.

49 F. Hu, T. K. Sarkar, and Y. Hua, "The Spectral Parameter Estimation by Using Prefiltering and Matrix Pencil Method," *Fifth IEEE ASSP Workshop on Spectrum Estimation and Modeling*, pp. 45–49, October 10-12, 1990.

50 T. Hua and T. K. Sarkar, "On SVD for Estimating Generalized Eigenvalues of Singular Matrix Pencil in Noise," *IEEE Trans. Signal Processing*, vol. 39, no. 4, pp. 892–900, April, 1991.

51 A. Swami, G. B. Giannakis, and J. M. Mendel, "Cumulant Based Approach to the Harmonic Retrieval Problem," *Proc. Intnl. Conf. Acoust., Speech, and Signal Processing*, New York, NY, pp. 2264–2267, April, 1988.

52 C. K. Papadopoulos and C. L. Nikias, "Parameter Estimation of Exponentially Damped Sinusoids Using Higher Order Statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 8, pp. 1424–1436, August, 1990.

53 Y. T. Chan, J. M. M. Lavoie, and J. B. Plant, "A Parameter Estimation Approach to Estimation of Frequencies of Sinusoids," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-29, no. 2, pp. 214–219, April, 1981.

54 B. Friedlander, "A Recursive Maximum Likelihood Algorithm for ARMA Spectral Estimation," *IEEE Trans. Info. Theory*, vol. IT-28, no. 4, pp. 639–646, April, 1981.

55 M. R. Matausek, S. S. Stankovic, and D. V. Radovic, "Iterative Inverse Filtering Approach to the Estimation of Frequencies of Noisy Sinusoids," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-31, no. 6, pp. 1456–1463, December, 1983.

56 S. S. Stankovic, M. Dragosevic, and M. Carapic, "Estimation of Frequencies of Sinusoids Using the Extended Kalman Filter," *IEEE Int. Conf. on Acoust., Speech, Signal Processing*, vol. III, paper 5.11/1–3, March, 1984.

57 K. S. Arun and M. Aung, "Detection and Tracking of Superimposed Non-Stationary Harmonics ," *Fifth IEEE ASSP Workshop on Spectrum Estimation and Modeling*, pp. 173–177, October 10-12, 1990.

58 A. Fredriksen, D. Middleton, and D. Vanderlinde,, "Simultaneous Signal detection and Estimation Under Multiple Hypotheses," *IEEE Trans. Info. Theory*, vol. IT-18, no. 5, pp. 607–614, September, 1972.

59 D. G. Lainiotis, "Optimal Adaptive Estimation : Structure and Parameter Adaptation," *IEEE Trans. Automatic Control*, vol. AC-16, no. 2, pp. 160–170, April, 1971.

60 H. Akaike, "A New Look at Statistical Model Identification," *IEEE Trans. Automatic Control*, vol. AC-19, no. 6, pp. 716–723, December, 1974.

61 J. Rissanen, "A Universal Prior for the Integers and Estimation by Minimum Descriptor Length," *Ann. Statist.*, vol. 11, pp. 416–431, 1983.

62 B. Baygun and A. O. Hero,, "An Order Selection Criteria via Simultaneous Estimation/Detection Theory," *Fifth IEEE ASSP Workshop on Spectrum Estimation and Modeling*, pp. 168–172, October 10-12, 1990.

63 J. Fuchs, "Estimating the Number of Sinusoids in Additive White Noise," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 36, no. 12, pp. 1846–1853, December, 1988.

64 S. S. Rao and C. Vaidyanathan, "Estimating the Number of Sinusoids in Non-Gaussian Noise Using Cumulants," *Proc. Intnl. Conf. Acoust., Speech, and Signal Processing*, paper 39.E7.11, May 14-17, 1991.

65 B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, 1979.

66 Knapp, C. H. and Carter, G. C., "The Generalized Correlation Method of Estimation of Time Delay, " *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-24, no. 4, pp. 320–327, August, 1976.

67 J. A. Stuller, "Maximum Likelihood Estimation of Time-Varying Delay – Part I," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-35, no. 3, pp. 300–313, March, 1987.

68 T. J. Abatzoglou, "A Fast Maximum Likelihood Estimator for Time Delay Estimation," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-34, no. 2, pp. 375–378, April, 1986.

69 R. E. Bethel and R. G. Rahikka, "Optimum Time Delay Detection and Tracking," *IEEE Trans. Aerosp. Elec. Syst.*, vol. 26, no. 5, pp. 700–711, September, 1990.

70 R. J. Kenefic, "A Bayesian Approach to Time Delay Estimation," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-29, no. 3, pp. 611–614, June, 1981.

71 G. C. Carter, "Time Delay Estimation for Passive Sonar Signal Processing," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-29, no. 3, pp. 463–470, June, 1981.

72 S. Stein, "Algorithms for Ambiguity Function Processing," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-29, no. 3, pp. 588–599, June, 1981.

73 C. L. Nikias and R. Pan, "Time Delay Estimation in Unknown Gaussian Spatially Colored Noise," *IEEE Trans. Acoust, Speech,and Signal Processing,* vol. ASSP-36, pp. 1706–1714, November, 1988.

74 H. H. Chiang and C. L. Nikias, "A New Method for Adaptive Time Delay Estimation," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-38, pp. 35–45, February, 1990.

75 A. D. Whalen, *Detection of Signals in Noise*, Academic Press, Inc., New York, 1971.

76 Glisson, T. H. and Sage, A. P., "On Sonar Signal Processing," *IEEE Trans. Aerosp. Elec. Syst.*, vol. AES-6, no. 1, pp. 37–50, January, 1970.

77 A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, 1984.

78 N. Levanon, *Radar Principles*, John Wiley and Sons, 1988.

79 H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part III Radar-Sonar Signal Processing and Gaussian Signals in Noise*, John Wiley and Sons, New York, 1971.

80 W. S. Burdic, *Underwater Acoustic System Analysis*, Prentice-Hall, Inc., 1984.

81 M. I. Skolnik, *Introduction to Radar Systems*, 2nd ed., McGraw-Hill, New York, 1980.

82 F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE,* vol. 66, no. 1, pp. 51–83, January, 1978.

Appendix

Cramer-Rao Bound for the Harmonic Retrieval Problem

This appendix presents the derivation of the Cramer-Rao bound for P exponentially damped sinusoids in white gaussian noise. Consider the measurement model given by the formula

$$\begin{aligned} z_k &= \sum_{p=1}^{P} c_{k_p} \exp\left(-\alpha_{k_p} k + j(\omega_{k_p} k + \theta_{k_p})\right) + v_k \\ &= h_k(\mathbf{x}_k) + v_k \end{aligned} \tag{A.1}$$

for $k = 0, 1, \cdots, K-1$. v_k is assumed to be complex white Gaussian noise with mutually independent real and imaginary components each with variance σ^2. The elements of the state variable vector $\mathbf{x}_k$ are defined as

$$\begin{aligned} x_{k_{4(p-1)+1}} &= \omega_{k_p} \\ x_{k_{4(p-1)+2}} &= c_{k_p} \\ x_{k_{4(p-1)+3}} &= \theta_{k_p} \\ x_{k_{4(p-1)+4}} &= \alpha_{k_p}. \end{aligned} \tag{A.2}$$

The objective is to estimate some or all of the $4P$ possibly time-varying parameters in this system based on the measurements. The probability density function of the set of measurments $\mathbf{z} = [z_0, z_1, \cdots, z_{K-1}]^T$ conditioned on the unknown parameters $\mathbf{x}_k$ is given by

$$p(\mathbf{z}|\mathbf{x}_k) = (2\pi\sigma)^{-K} \exp\left[\frac{-1}{2\sigma^2}\sum_{k=0}^{K-1} |z_k - h_k(\mathbf{x}_k)|^2\right] \qquad (A.3)$$

The Cramer-Rao bound [73] (pp. 66, 84) gives the minimum possible variance of any unbiased estimate $\hat{\mathbf{x}}_k(\mathbf{z})$ of the state $\mathbf{x}_k$. In the presence of no prior information about the state this bound is given by

$$\mathrm{Var}[\hat{\mathbf{x}}_k(\mathbf{z}) - \mathbf{x}_k] = J_D^{-1} \qquad (A.4)$$

where J_D is the Fisher information matrix given by

$$J_D = -E\left[\frac{\partial^2 \ln p(\mathbf{z}|\mathbf{x}_k)}{\partial \mathbf{x}_k \partial \mathbf{x}_k^T}\right]$$

Applying this to (A.3) yields

$$J_D = \frac{1}{2\sigma^2}\sum_{k=0}^{K-1} E\left\{\frac{\partial^2}{\partial \mathbf{x}_k \partial \mathbf{x}_k^T}(z_k\, z_k^* - z_k\, h_k(\mathbf{x}_k)^* - z_k^*\, h_k(\mathbf{x}_k) + h_k(\mathbf{x}_k)\, h_k(\mathbf{x}_k)^*)\right\}$$

which reduces to

$$J_D = \frac{1}{2\sigma^2}\sum_{k=0}^{K-1} E\left\{\frac{\partial h_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}\,\frac{\partial h_k(\mathbf{x}_k)^*}{\partial \mathbf{x}_k^T} + \frac{\partial h_k(\mathbf{x}_k)^*}{\partial \mathbf{x}_k}\,\frac{\partial h_k(\mathbf{x}_k)}{\partial \mathbf{x}_k^T}\right\} \qquad (A.5)$$

Noting that $\frac{\partial h_k(\mathbf{x}_k)}{\partial \mathbf{x}_k^T} = \left(\frac{\partial h_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right)^T$, (A.5) can be evaluated by finding expressions for $\frac{\partial h_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}$ and $\frac{\partial h_k(\mathbf{x}_k)^*}{\partial \mathbf{x}_k}$ .Let

$$h_k(\mathbf{x}_k) = \sum_{p=1}^{P} h_{k_p}(\mathbf{x}_k)$$

where the measurement component $h_{k_p}(\mathbf{x}_k)$ represents the contribution from the p^{th} sinusoid and is given by

$$h_{k_p}(\mathbf{x}_k) = c_{k_p} \exp\left(-\alpha_{k_p} k + j(\omega_{k_p} k + \theta_{k_p})\right)$$

The two partial derivative vectors can then be expressed as

$$\begin{aligned} \frac{\partial h_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} &= \sum_{p=1}^{P} \begin{bmatrix} jk \\ 1/c_{k_p} \\ j \\ -k \end{bmatrix} h_{k_p}(\mathbf{x}_k) \\ \frac{\partial h_k(\mathbf{x}_k)^*}{\partial \mathbf{x}_k} &= \sum_{p=1}^{P} \begin{bmatrix} -jk \\ 1/c_{k_p} \\ -j \\ -k \end{bmatrix} h_{k_p}(\mathbf{x}_k)^* \end{aligned} \tag{A.6}$$

A.1 Case 1 : Estimation of two Frequencies and Damping Coefficients

For estimation of the parameters of two exponentially damped sinusoids the measurement equation becomes

$$z_k = \sum_{p=1}^{P} h_{k_p}(\mathbf{x}_k) \; + \; v_k$$

where

$$h_{k_p}(\mathbf{x}_k) = \exp\left(-\alpha_{k_p} k + j\omega_{k_p} k\right) \tag{A.7}$$

The elements of the state variable vector $\mathbf{x}_k$ are defined as

$$\begin{aligned} x_{k_1} &= \omega_1 \\ x_{k_2} &= \alpha_1 \\ x_{k_3} &= \omega_2 \\ x_{k_4} &= \alpha_2. \end{aligned} \tag{A.8}$$

and the partial derivatives in (A.5) become

$$\frac{\partial h_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} = k \begin{bmatrix} j\, h_{k_1}(\mathbf{x}_k) \\ -h_{k_1}(\mathbf{x}_k) \\ j\, h_{k_2}(\mathbf{x}_k) \\ -h_{k_2}(\mathbf{x}_k) \end{bmatrix} \quad \frac{\partial h_k(\mathbf{x}_k)^*}{\partial \mathbf{x}_k} = k \begin{bmatrix} -j\, h_{k_1}(\mathbf{x}_k)^* \\ -h_{k_1}(\mathbf{x}_k)^* \\ -j\, h_{k_2}(\mathbf{x}_k)^* \\ -h_{k_2}(\mathbf{x}_k)^* \end{bmatrix}. \tag{A.9}$$

The Fisher information matrix for this system becomes

$$J_D = \frac{1}{\sigma^2} \begin{bmatrix} a & 0 & c & 0 \\ 0 & a & 0 & c \\ c & 0 & b & 0 \\ 0 & c & 0 & b \end{bmatrix} \tag{A.10}$$

where

$$a = \sum_{k=0}^{K-1} k^2\, e^{-2\alpha_1 k}$$

$$b = \sum_{k=0}^{K-1} k^2\, e^{-2\alpha_2 k}$$

$$c = \sum_{k=0}^{K-1} k^2\, e^{-(\alpha_1+\alpha_2) k} \cos((\omega_1 - \omega_2)\, k)$$

Inverting this analytically the CR bound becomes

$$\mathrm{Var}[\hat{\mathbf{x}}_k(\mathbf{z}) - \mathbf{x}_k] = \sigma^2 \begin{bmatrix} a' & 0 & c' & 0 \\ 0 & a' & 0 & c' \\ c' & 0 & b' & 0 \\ 0 & c' & 0 & b' \end{bmatrix}$$

where

$$a' = \frac{a\,b^2 - b\,c^2}{a^2\,b^2 - 2\,a\,b\,c^2}$$

$$b' = \frac{a^2\,b - a\,c^2}{a^2\,b^2 - 2\,a\,b\,c^2}$$

$$c' = \frac{c(a\,b - c^2)}{a^2\,b^2 - 2\,a\,b\,c^2}$$

From this the following observations are made:

1 For a given sinusoid, the CR bound for the frequency estimate is the same as that for the damping coefficient estimate.

2 A higher damping coefficient gives a larger CR bound.

3 The bound is dependent on the difference between the two frequencies and not their individual values.

A.2 Case 2 : CR Bound with *a priori* Information

In the case where the statistics of the initial estimation error are known and are given by the initial covariance P_0 the CR bound (A.4) becomes

$$\mathrm{Var}[\hat{\mathbf{x}}_k(\mathbf{z}) - \mathbf{x}_k] = [J_D + J_P]^{-1}. \qquad (A.11)$$

where $J_P = 1/P_0$.

If J_D is small relative to J_P then the bound will be controlled by the initial covariance. From equation (A.5) it can be seen that as the noise variance σ^2 increases J_D decreases. However, even though the increased noise increases the CR bound, there will be a point when the J_P dominates implying that the optimal unbiased estimate is always at least as good as the initial estimate. Thus J_P sets an upper limit on the variance in the estimate.